UNIVERSITY
OF SUSSEX

Doctoral Thesis

# Cosmological information from moments and Bayesian reconstruction methods

*A thesis submitted in fulfilment of the requirements*
*for the degree of Doctor of Philosophy*

*in the*

Department of Physics & Astronomy
Faculty of Science

*Author:*
Lucas Porth

*Supervisor:*
Robert E. Smith

1st July 2021

# Declaration

I, Lucas Porth, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

Part of this thesis has been completed in collaboration with Robert E. Smith, Gary M. Bernstein, Patrick Simon, Laura Marian and Stefan Hilbert, and is comprised of the following publications:

- L. Porth, R. E. Smith, P. Simon, L. Marian, S. Hilbert, Fast estimation of aperture-mass statistics I: aperture mass variance and an application to the CFHTLenS data, published in MNRAS $499(2) : 2474 - 2492$
  Robert E. Smith came up with the original idea of this paper. The calculations, implementation and analysis were completed by myself under the supervision of Robert. E. Smith. The shape catalog of the CFHTLenS data was provided by Patrick Simon. The ray-tracing simulations were performed by Stefan Hilbert from which the mock catalogs were constructed by Laura Marian and provided by Robert E. Smith.

- L. Porth, R. E. Smith, Fast estimation of aperture mass statistics II: Detectability of higher order statistics in current and future surveys, arXiv:2106.04594, submitted to MNRAS.
  Robert E. Smith came up with the original idea of this paper which was then refined and generalized by myself and Robert E. Smith. The calculations, implementation and analysis were completed by myself under the supervision of Robert E. Smith. In this thesis we additionally include an appendix (3E) in which we show the sampling distribution of the estimators.

- L. Porth, G. M. Bernstein, R. E. Smith, The information content of projected galaxy

fields, to be submitted.

Gary M. Bernstein came up with the original idea of this project. Additionally, he suggested the structure of the model described in section 4.2. The implementation and analysis was completed by myself under the supervision of Gary M. Bernstein.

*Brighton,*

*1st July 2021*

Lucas Porth

UNIVERSITY OF SUSSEX

Lucas Porth, Doctor of Philosophy

# Cosmological information from moments and Bayesian reconstruction methods

## Summary

This thesis explores novel ways on constraining the cosmological model of our Universe with observational data. The explosion in high fidelity data coupled with the increasing number of complementary cosmological probes, makes it possible to test our theories to an unprecedented precision. However, reaching the experimental designed precision for the upcoming surveys will require advances in theoretical and statistical modelling. We contribute work to two avenues addressing this problem, in particular by extracting nongaussian information from higher order weak gravitational lensing statistics and by using Bayesian forward modelling approaches to increase the information gain compared to traditional analysis methods.

For the first project we focus on the aperture mass statistics, which is a E/B decomposed measure of the weak lensing convergence polyspectra. We construct an accelerated estimator that will make it feasible to measure those statistics in current and forthcoming surveys. After identifying an optimized weighting scheme and adapting the theoretical expressions to account for potential biases, we successfully test the estimators' performance and reliability on simulated surveys and then measure the second order statistics on the CFHTLenS survey, finding excellent agreement with the traditional analysis in terms of the shear correlation functions.

For the second project we move away from summary statistics and shift our focus to the cosmic field itself. Using idealized structure formation models we construct a likelihood function for a linearized cosmic field, taking into account the observed realization of tracers. For an efficient sampling of this high dimensional problem we employ a Hamiltonian Monte Carlo framework. We then extend the likelihood to jointly sample the field and the amplitude of its underlying power spectrum. We find that for realistic scale cuts and galaxy number densities this framework increases the information content by a factor of four.

# Acknowledgements

First of all, I would like to thank my supervisor Robert Smith for bringing me to Brighton and for guiding me through the four years of PhD experience. Thank you very much, Robert, for all the good scientific and general advice, for making sure that the project stays on track, and for helping me to develop a more critical scientific thinking. It's been a fun ride!

I also owe a big thanks to Gary Bernstein who acted as another supervisor since my stay in Philly and who introduced me to the world of high dimensional Bayesian statistics. Thank you for giving me the opportunity to work with you and for all your clear explanations that always made me leave your office much less confused than I entered it.

I would also like to thank Jon Loveday who acted as my second supervisor during the first bit of the PhD and helped me a lot to understand and analyze "real life" data.

My time in Brighton would not have been the same without all the amazing people I had the privilege to meet - in totally random order: Chris, Aswin, Ciaran, Raph, Steven, Hannah, Michaela, Margherita, Azizah, Ridwan, Jussi, Dan, Michele, Jesus, Daniel, Luke, Rose, Stephen, Mark, David, Kareem, Itzi, Julien, Mateja, Ze, Alex, Ian, Edward, Will, Robin, Pippa, T-Fi, Antony, Maria, Benoit, Reese, Cameron, Chris, Dimitrios, Edward, Giulio, Andrew, Carlos and Sunayana. Thanks for the lunches, beers at Falmer and beyond, competitive squash and badmingon sessions, as well as all the other non-academic activities that will definitely remain as positive memories. In particular, I would like to thank my office mates: Alex and Ridwan for their warm welcome and help with so many little things, Dan for being a great mate, discussion partner and company during the whole PhD life, Cameron for the banter, Stephen for not(!) always being in before me and Robin for never saying no to Friday lunches.

I would also like to thank the crew at Penn for making my stay with its tennis and ping-pong sessions, axe-throwing and rooftop bourbonades so worthwhile - again, in randomized order: Dillon, Lucas II, Tae, Mariana, Pedro, Eric, Shivam, Anu, Guram, Tasha, Ling and Cyrille.

For maintaining a bit of sanity during the weird times in 2020 and 2021 I consider myself extremely fortunate to have Aswin and David as flatmates. Thanks dudes for the good chats + banter, the random food, the rum diaries, and all the groovy Fridays!

Last, but definitely not least, I want to thank my parents. Thank you so much for your unwavering support and belief in me! Without your help and guidance throughout all the time I would not be able to stand where I am now.

# Contents

# List of Figures

# Introduction

# 1

In this chapter we give an overview of the main ideas that are needed to build a simplified and predictive model of our Universe. In §1.1 we begin by laying out the geometric foundation of an expanding spacetime and explore the equations governing a universe's dynamics given its contents. After collecting various distance measures, in §1.2 we continue by refining the cosmological model to include small perturbations that will evolve and may finally collapse into the virialized structures we observe today. We then introduce various summary statistics of the density field that may be obtained from both, theory and observations and give a brief overview of theoretical challenges for their accurate prediction. In §1.3 we introduce the theory of weak gravitational lensing and motivate some key equations that will be used in the later chapters of this thesis. Finally, in §1.4 we introduce the Bayes paradigm in a scientific context and give an overview of Markov Chain Monte Carlo (MCMC) methods.

The content of this chapter can be found in various textbooks on general relativity and cosmology, as well as in more recent review articles. In the first two sections we mainly follow along the lines of Dodelson (2003), Wald (1984), Piattella (2018) and Lesgourgues & Pastor (2006), while for §1.3 we made use of Bartelmann & Schneider (2001), Kilbinger (2015) and Mandelbaum (2018) and §1.4 borrowed information found in Heavens (2009). Throughout the chapter we add additional references pointing to the original literature.

## 1.1 The expanding Universe

### 1.1.1 Foundations

The standard model of relativistic cosmology is grounded on three assumptions: General relativity (GR), the cosmological principle, and Weyl's postulate. The first assumptions equips us with a mathematical formalism to build upon; in particular we will define a universe as a four dimensional spacetime obeying the Einstein field equations (Einstein, 1915)

$$G_{\alpha\beta} \equiv R_{\alpha\beta} - \frac{1}{2} R g_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta} \ , \tag{1.1}$$

which link the geometry of spacetime encapsulated in the Einstein tensor $G_{\alpha\beta}$ to its content described by the energy-momentum tensor $T_{\alpha\beta}$. The Einstein tensor itself depends on the Ricci tensor $R_{\alpha\beta}$ and its contraction, the Ricci scalar $R$. All those geometric quantities in turn can be constructed from the second derivatives of the more fundamental metric tensor $g_{\alpha\beta}$, which itself can be used to measure distances in spacetime.

The idea of the cosmological principle dates back much further than the latter two assumptions - a first conceptual formulation can already be found in Newton's Principia Mathematica, in which he proposes a universal law of gravitation that extends the equations of motion through absolute space. Phrased in a modern jargon, the cosmological principle asserts homogeneity and isotropy on spacelike hypersurfaces $\Sigma_t$, i.e. that at each point in time the universe looks the same in each direction for each point in space. The formal definitions of spatial homogeneity and isotropy then imply that there exists a foliation of spacetime into spacelike hypersurfaces, that the $\Sigma_t$ are orthogonal to an observers (timelike) worldline and that each such $\Sigma_t$ has to be a space of constant curvature. The latter property in turn reduces the number of each $\Sigma_t$s possible geometries to three; namely to a spherical, flat, or hyperbolical one. For this thesis we will exclusively work in the flat case, which is supported by recent observations (Alam et al., 2017; Efstathiou & Gratton, 2019). The resulting form of the metric is known as the Friedman-Lemaitre-Robertson-Walker (FLRW) metric (Friedman, A., 1922; Lemaître, 1931; Robertson, 1935; Walker, 1937):

$$\mathrm{d}s^2 = -c^2 \mathrm{d}t^2 + a^2(t) \left[ \mathrm{d}\chi^2 + \chi^2 \mathrm{d}\Omega^2 \right] \ , \tag{1.2}$$

where the positive function $a(t)$ denotes the scale factor describing the expansion of the universe with time $t$ and the spatial part of the metric is expressed in spherical coordinates consisting of the comoving distance $\chi$ and the solid angle $\Omega$ . As this line element is invariant under the scaling transformation $(a \to \kappa a \ , \ \chi \to \chi/\kappa)$ one can set the scale factor today to unity, $a_0 \equiv 1$.

Finally, Weyl's postulate (Weyl, 1923) states that the cosmic fluid particles move along non intersecting timelike geodesics originating from a single point. The postulate implies that observers following different world lines can agree on a common, universal time. One should note that there is some debate on whether Weyl's postulate is a necessary assumption or if it can be absorbed in the cosmological principle, see for example Rugh & Zinkernagel (2010) and references therein.

From the time component of the geodesic equation for a massless particle (i.e. a photon) propagating through an FLRW spacetime one deduces that the particles energy is inversely proportional to the scale factor. This implies that if we observe a photon with wavelength $\lambda_{\rm obs}$ its wavelength at emission was different. We denote this difference as the redshift $z \equiv (\lambda_{\rm obs} - \lambda_{\rm em})/\lambda_{\rm em}$. Using that a photons energy $E$ is also proportional to its frequency $f$ we can relate the redshift to the scale factor as:

$$a_{\rm em} = \frac{E_{\rm obs}}{E_{\rm em}} = \frac{f_{\rm obs}}{f_{\rm em}} = \frac{\lambda_{\rm em}}{\lambda_{\rm obs}} = \frac{1}{1+z_{\rm em}} \,, \tag{1.3}$$

where we assumed the observing time to be today.

### 1.1.2 Cosmic Dynamics

With the FLRW metric in hand one can express the left hand side of Eq. (1.1) in terms of the scale factor and its derivatives " $\cdot$ " with respect to time. For finding an explicit expression for the energy-momentum tensor we require it to have the same symmetries as Eq. (1.2); Weinberg (1972) shows that this implies the form of an ideal fluid

$$T_{\alpha\beta} = \left(\rho c^2 + p\right) u_\alpha u_\beta - p\, g_{\alpha\beta} \,, \tag{1.4}$$

where $\rho$, $p$ and $u$ denote the fluid particles' energy density, pressure and 4-velocity. Due to the orthogonality of the fluids' worldines to the spatial part of the metric, the velocities in the comoving frame are simply given by $u \sim (ca, 0, 0, 0)$. Subbing those expression in the Einstein equations we arrive at the two Friedmann equations describing the dynamical evolution of a homogeneous universe:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \tag{1.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{p}{c^2}\right) \,. \tag{1.6}$$

We see that expansion rate of the universe Eq. (1.5) is governed by its total energy density, while it's acceleration Eq. (1.6) does additionally depend on the fluid's pressure[1]. The expansion rate

[1] Had we not restricted ourselves to the spatially flat case, Eq. (1.5) would also depend on the universes' global geometry.

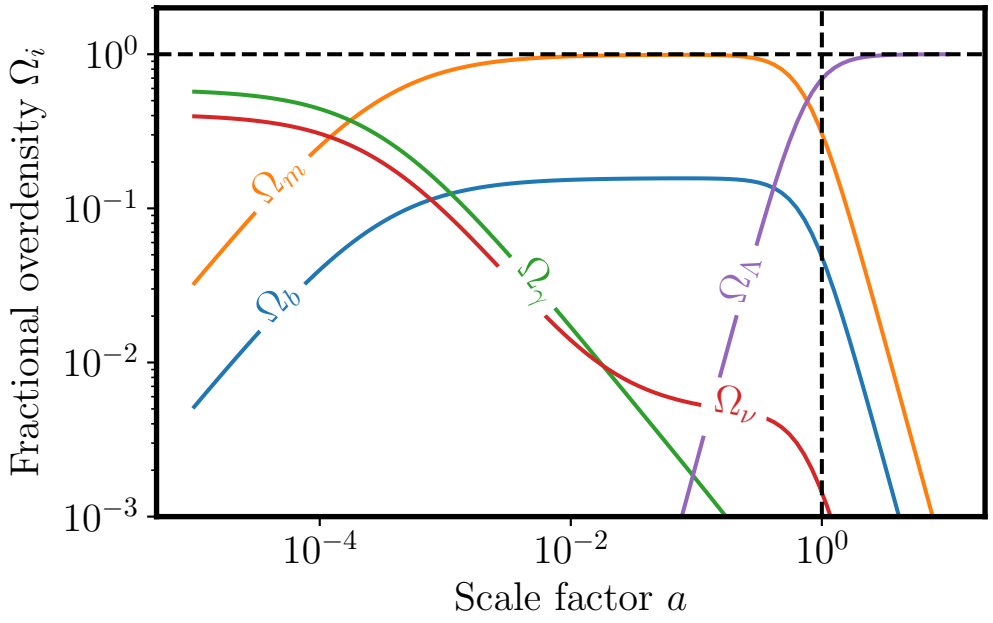


Figure 1.1 Evolution of the different energy components of the Universe. At each time the components add up to unity as indicated by the horizontal dashed line. The present is indicated by the vertical dashed line - the intersections of the individual components with that line correspond to the values or the Planck Collaboration et al. (2020) analysis.

of the universe can equivalently labeled by the Hubble factor $H \equiv \dot{a}/a$.

From the covariant conservation of the energy-momentum tensor one obtains the continuity and Euler equations, the former is given by

$$\dot{\rho} + 3H\left(\rho + \frac{p}{c^2}\right) = 0 \quad \Rightarrow \quad \rho_i(a) = \rho_{i,0} \exp\left[3\int_a^1 \frac{1+w_i(a')}{a'}\mathrm{d}a'\right], \tag{1.7}$$

where we solved for an equation of state, $p_i \equiv w_i(a)\rho_i c^2$ of some energy component $i$. To arrive at a more intuitive form we define the dimensionless fractional density parameters $\Omega_i \equiv \rho_i/\rho_{\mathrm{crit.}} \equiv 8\pi G\rho H_0^{-2}/3$ in terms of the critical density $\rho_{\mathrm{crit.}}$. For a universe resembling the matter content of our Cosmos the first Friedmann equation Eq. (1.5) can now be rewritten as

$$\left(\frac{H}{H_0}\right)^2 = \Omega_\Lambda + \frac{\Omega_{\mathrm{m},0}}{a^3} + \frac{\Omega_{\mathrm{b},0}}{a^3} + \frac{\Omega_{\mathrm{r},0}}{a^4} + \Omega_{\nu,0}\exp\left[3\int_a^1 \frac{1+w_\nu(a')}{a'}\mathrm{d}a'\right], \tag{1.8}$$

where the energy content is distributed between non-relativistic matter ($w = 0$) consisting of baryons $\Omega_{\mathrm{b}}$, cold dark matter $\Omega_{\mathrm{c}}$ and massive neutrinos, a relativistic contribution $\Omega_{\mathrm{r}}$ ($w = -1/3$) made up by photons and massless neutrinos, as well as a cosmological constant $\Omega_\Lambda$ ($w = -1$). Evaluating Eq. (1.8) at the present time we see that the fractional densities add up to unity. One big aim of observational cosmology is to constrain these parameter as tight as possible - for this introduction we adapt all numerical calculations to the match the recent constraints from the Planck mission (Planck Collaboration et al., 2020).

From the $a$-dependence in Eq. (1.8) we can deduce that the different energy components will dominate the universe at different times, the numerical solution is shown in Fig. 1.1. In the early stages the relativistic components dominate, followed by a phase of matter domination, until finally the cosmological constant is taking over to dominate the energy content from the universe for the rest of its existence. The curve for the neutrino density is the only one to have an inflection point as massive neutrinos transition from being a relativistic species at early times to a matter like component at later times; the location of the inflection point does in particular depend on the neutrino mass scale.

### 1.1.3 Distances and horizons

Due to the dynamic geometry of the FLRW metric there does not exist a universal way to measure distances. The most prominent quantity is the comoving distance $\chi$ being already defined in the FLRW metric, one can relate it to the observable redshift as

$$\chi(z) \equiv \int_{t(z)}^{t_0} c\frac{\mathrm{d}t}{a(t)} = \int_{a(z)}^{1} c\frac{\mathrm{d}a'}{a'^2 H(a')} = \int_0^z c\frac{\mathrm{d}z'}{H(z')} \,. \tag{1.9}$$

Note that $\chi$ remains constant for observers moving with the Hubble flow (Hogg, 1999) - for obtaining the physical (proper) distance $\chi_{\mathrm{prop}}$ light is travelling between two observers one needs to take into account the time evolution of the expansion factor:

$$\chi_{\mathrm{prop}}(z) \equiv \int_{a(z)}^{1} c\frac{\mathrm{d}a'}{a' H(a')} = \int_0^z c\frac{\mathrm{d}z'}{(1+z')H(z')} \,. \tag{1.10}$$

Neither the comoving nor the proper distance are directly observable, however there exist two additional measures that are both, linkable to observations and directly related to $\chi$. The first one is the angular diameter distance $\chi_A$, which is defined as the ratio of an objects size (perpendicular to the line of sight) and its angular size

$$\chi_A(z) \equiv \frac{\chi(z)}{1+z} \,. \tag{1.11}$$

Quantities or features with a known absolute size are called standard rulers. By measuring their apparent scales one can infer the distance to the observer using Eq. (1.11). On the other hand, if an object is a standard candle (i.e. has a known absolute luminosity $L$) one can infer the comoving distance by virtue of the luminosity distance

$$\chi_L(z) \equiv (1+z)\chi(z) \,, \tag{1.12}$$

which is constructed to obey the relation $F = L/(4\pi\chi_L^2)$ for the observed flux $F$. We collect all distance measures in Fig. 1.2.

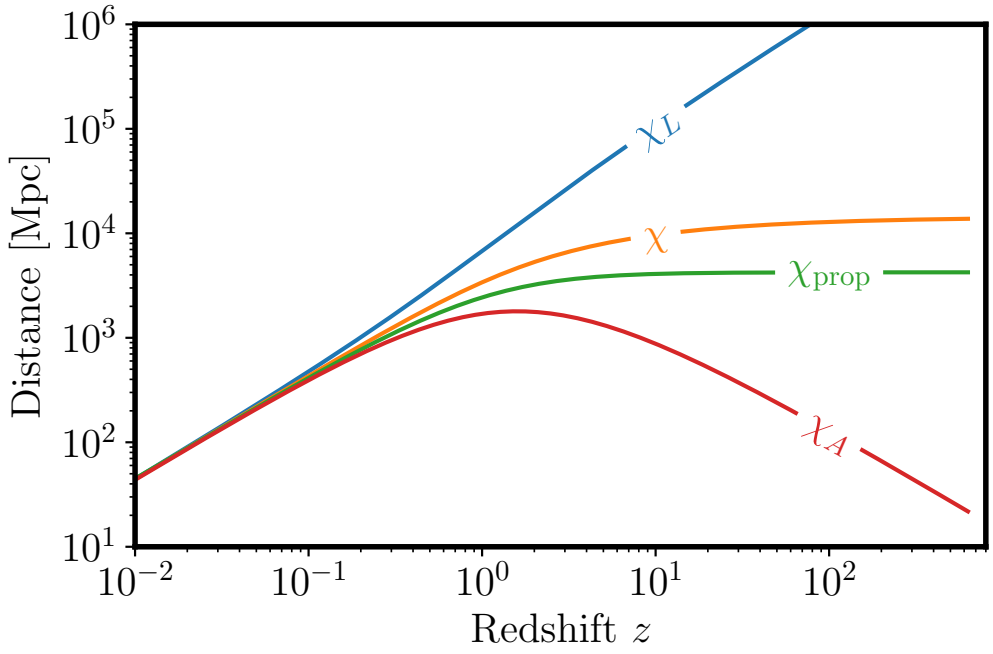


Figure 1.2 Evolution of the various distance measures with cosmological redshift. Note the turnover point of the angular diameter distance implying that objects appear magnified for large redshifts.

## 1.2 Structure formation

Our nearby universe is far from homogeneous on all scales; instead we observe highly overdense galaxies that appear to trace a more continuous underlying cosmic web. From our assumption of GR and the structure seen today one might hypothesize that today's observations could have evolved from small departures of homogeneity at an earlier stage of the Cosmos.

### 1.2.1 The early universe

For this thesis we assume that there exists a mechanism that gave rise to some cosmic initial conditions that are consistent with the current observations of the Cosmic Microwave Background (CMB). The most widely accepted models are based on the inflationary paradigm (Kazanas, 1980; Guth, 1981; Linde, 1982; Albrecht & Steinhardt, 1982), in which additional scalar field(s) source a close to exponential expansion of the cosmos prior to radiation domination. While inflationary models were primarily put forward to find explanations for the flatness, horizon and magnetic monopole problems[2], it was realized by Mukhanov & Chibisov (1981) that quantum fluctuations

[2] Note however the criticism of Penrose (1989) estimating the amount of fine-tuning necessary for inflation to occur to be much larger than the one it is able to resolve.

within the inflaton field will expand to cosmological scales and can serve as the primordial seeds (initial conditions) for the large-scale structure we observe today.

### 1.2.2 Prosaic perturbation theory

Let us now proceed by modelling a line element that is capable of including the primordial fluctuations. This can formally be achieved by adding perturbations to the FLRW metric (1.2):

$$\mathrm{d}s^2 = c^2 a^2(\tau)\left[-(1+2A)\mathrm{d}\tau^2 + 2B_i \mathrm{d}\tau \mathrm{d}x^i - (2C\delta^K_{ij} + D_{ij})\mathrm{d}x^i \mathrm{d}x^j\right] , \tag{1.13}$$

where we introduced the conformal time $\mathrm{d}\tau \equiv \mathrm{d}t/a(t)$. A suitable reparametrization can be obtained by the so-called SVT decomposition, in which the perturbation quantities are decomposed in their scalar, vector and tensor components (Lifshitz, 1946), giving four scalar, two (divergenceless) vector and a (transverse and traceless) tensorial variable. One can show that due to the homogenous and isotropic background these quantities evolve independently in the (first order) perturbed Einstein tensor, (e.g. Durrer (2008)). To take care of spurious perturbations stemming from a specific coordinate choice one first computes the transformation equations of the perturbation variables under a general coordinate (gauge) transformation and from there constructs new, gauge invariant, quantities that are coordinate independent and therefore reflect the physical degrees of freedom (Bardeen, 1980). This procedure eliminates four components (two scalar and a vector one), just as one would have expected from the diffeomorphism invariance of GR. For this thesis we will not work in a fully gauge invariant fashion, but instead fix the coordinate system to the so-called conformal Newtonian gauge (Mukhanov et al., 1992) and will furthermore restrict ourselves to scalar perturbations; due to this choice the corresponding line element remains diagonal:

$$\mathrm{d}s^2 = a^2(\tau)\left[(1+2\psi)\mathrm{d}\tau^2 - (1-2\phi)\mathrm{d}x_i \mathrm{d}x^i\right] . \tag{1.14}$$

The two remaining perturbation variables are commonly referred to as the Bardeen potentials. Plugging this line element in the Einstein equations and keeping only first order terms one gets the perturbed Einstein tensor. For the corresponding right hand side one constructs a perturbed energy-momentum tensor in a similar fashion, introducing the density perturbation $\delta\rho$, the pressure perturbation $\delta p$, as well as the velocity perturbation $v^i$ and the anisotropic stress $\sigma$. Both sides together yield the perturbed Einstein equations; four equations of motion for the two metric perturbation potentials in terms of the four scalar energy component perturbations. Structurally they read:

$$\delta G_{\alpha\beta} = 4\pi G a^2 \sum_i \delta T_{\alpha\beta,i} , \tag{1.15}$$

where the sum runs over different energy components.

Getting the corresponding equations of motion for the right hand side of (1.15) turns out to be a more complicated task. From the local covariant conservation of the energy momentum tensor one can again obtain a continuity and an Euler equation for the perturbation variables. For some energy component $i$ they read in Fourier space (Ma & Bertschinger, 1995):

$$\delta_i' = -(1+w_i)(\theta_i - 3\phi') - 3\mathcal{H}\left(c_{s,i}^2 - w_i\right)\delta_i \tag{1.16}$$

$$\theta_i' = -\mathcal{H}(1-3w_i)\theta_i - \frac{w_i'}{1+w_i}\theta_i + \frac{c_{s,i}^2}{1+w_i}k^2\delta_i - k^2\sigma_i + k^2\psi\,, \tag{1.17}$$

where we defined the overdensity $\delta \equiv \delta\rho/\rho - 1$, the speed of sound $c_s^2 \equiv \delta P/\delta\rho$ and the velocity divergence $\theta \equiv k_i v^i$ and denoted derivatives with respect to conformal time as $'$ and the Hubble factor expressed through $\tau$ as $\mathcal{H}$. In case of a non-interacting species with vanishing anisotropic stress and pressure perturbations (such as dark matter) the system of equations is closed, otherwise we need to find some more equations of motion. This can be achieved when switching to a microscopic definition of the energy-momentum tensor in terms of the particle distribution function $f$ in phase space that in turn is governed by the Boltzmann equation:

$$\frac{\partial f}{\partial \tau} + \frac{\mathrm{d}x^i}{\mathrm{d}\tau}\frac{\partial f}{\partial x^i} + \frac{\mathrm{d}q}{\mathrm{d}\tau}\frac{\partial f}{\partial q} + \frac{\mathrm{d}n^i}{\mathrm{d}\tau}\frac{\partial f}{\partial n^i} = \left(\frac{\partial f}{\partial \tau}\right)_{\text{coll.}}\,, \tag{1.18}$$

where the phase space is spanned by the positions $x_i$, as well as the modulus $q \equiv ap$ and direction $\hat{n}_i \equiv q^i/q$ of the three-momentum. Perturbing the distribution function as $f(x^i, p^j, \tau) \to f_0(q)\left[1 + \Psi\left(x^i, q, n^j, \tau\right)\right]$ and using the metric (1.14) in (1.18) one derives an equation of motion for $\Psi$. Finally, expanding $\Psi$ in terms of Legendre polynominals one can carry out the integrations yielding an infinite hierarchy of equations for the expansion coefficients $\Psi_\ell$. For a more detailed account on relativistic cosmological perturbation theory see i.e. (Ma & Bertschinger, 1995).

Efficiently solving the set of coupled equations (1.15)-(1.17) together with the various hierarchies one requires multiple approximation schemes which a discussion of is beyond the scope of this thesis. Fortunately there exist well maintained public implementations for that problem; the most well known ones are CAMB and CLASS (Lewis et al., 2000; Blas et al., 2011), which themselves build on the earlier codes CMBFAST and COSMICS (Seljak & Zaldarriaga, 1996; Bertschinger, 1995).

In Fig. 1.3 we show the time evolution of the gravitational potentials and the overdensity parameters for our toy cosmology as calculated by CLASS. We see that on superhorizon scales the two metric potentials are frozen, but not equal to each other - the latter point is due to the fact that in that regime free-streaming (relativistic) neutrinos have a nonvanishing anisotropic stress component. Small scale perturbations will re-enter the horizon during radiation domination where they become strongly suppressed. While during matter domination the perturbations are

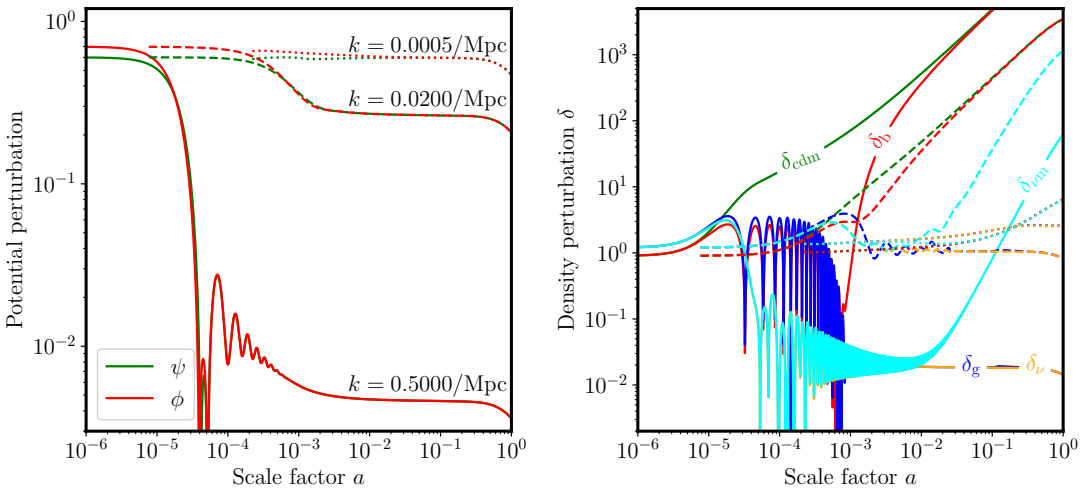


Figure 1.3 *Left hand side*: Evolution of Bardeen potentials in linear perturbation theory. *Right hand side*: Evolution of the fractional overdensities of various energy components. The linestyles are matched to the plot on the left

roughly frozen and equal to each other they start decreasing again once the cosmological constant takes over. Multiple interesting effects can also be observed when tracing the evolution of the individual overdensity parameters:

- During radiation domination superhorizon scale dark matter perturbations grow as $a^2$ whereas upon horizon entry the growth gets suppressed to follow a logarithmic scaling (Meszaros, 1974). Moving to the matter dominated era, the perturbations grow as the scale factor as long as the neutrinos remain relativistic; at later stages the exponent is slightly suppressed due to free-streaming massive neutrinos (Bond et al., 1980).

- Before recombination photons and baryons are tightly coupled via Thompson scattering and the they undergo acoustic oscillations (Peebles & Yu, 1970). For small scales the oscillation amplitude is strongly suppressed due to photon diffusion (Silk, 1968). After recombination the baryons fall into the gravitational potentials of dark matter such that their perturbations will converge to the dark matter ones. In contrast, the photons remain in a free streaming state such that their perturbations are frozen.

- In the radiation dominated era the neutrinos are relativistic and free-stream such that their perturbations are damped on subhorizon scales. During matter domination it is the neutrino mass and its associated scale factor $a_{\rm nr}$ (specifying the time at which the non-relativistic transition occurs) that determine the subsequent evolution of the perturbations: while relativistic neutrinos evolve as $\delta_\gamma$, massive ones track $\delta_{\rm b}$. Therefore, large scales en-

tering the horizon after $a_{\rm nr}$ will track $\delta_{\rm cdm}$ while smaller scale perturbations start growing from gradually lower values compared to $\delta_{\rm cdm}$ such that they still remain suppressed up to now.

For a thorough justification of our qualitative discussion we refer to Lesgourgues & Pastor (2006).

### 1.2.3 Going nonlinear

As we have seen in Fig. 1.3 the gravitational potentials on subhorizon scales are both frozen and equal to each other during matter domination. We might interpret this nice feature to give us permission to use Newtonian mechanics to track the evolution of perturbations in this regime. If we now want to describe the nonlinear evolution of cold dark matter we can redo a Newtonian version of relativistic perturbation theory in which the continuity, Euler and Poisson equations become:

$$\dot{\delta} + \frac{1}{a}\nabla \cdot [(1+\delta)\boldsymbol{v}] = 0 \tag{1.19}$$

$$\dot{\boldsymbol{v}} + H\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\frac{1}{a}\nabla\phi \tag{1.20}$$

$$\Delta\phi = \frac{3\Omega_m H_0^2}{2a}\delta\,, \tag{1.21}$$

where $\delta \equiv \rho/\bar{\rho} - 1$ is the density contrast with respect to the global background density $\bar{\rho}$ and $\boldsymbol{v} \equiv \boldsymbol{u} - H\boldsymbol{x}$ denotes the peculiar velocity field, i.e. the relative velocity with respect to the Hubble flow. An early ansatz to analytically solve those equations is given by standard perturbation theory (SPT) (Juszkiewicz, 1981; Fry, 1984; Goroff et al., 1986; Jain & Bertschinger, 1994), in which the nonlinear dynamical quantities are equated to a power series of the linear ones. Applying the this ansatz in Fourier space one can then find a recursive solution for the nonlinear mode coupling kernels. While this framework gives a good description for high redshifts, it is inherently limited by the requirement of small overdensities and suffers from multiple convergence issues (Blas et al., 2011). To address some of these shortcomings there have been developed multiple further techniques, such at resummation schemes (Crocce et al., 2006; Bernardeau et al., 2008) and effective theories based on course grained versions of the fluid equations and symmetry considerations (Baumann, 2009; Carrasco et al., 2014). Another interesting observation has been made by Coles & Jones (1991), who showed that under the assumption of a Gaussian initial velocity field Eq. (1.19) is solved by a lognormal distribution for the density field $\rho$ - we will come back to this model in Chapter 4.

All analytic methods developed so far can only describe the density within the linear and parts of the quasilinear regime; for a proper treatment of smaller scales one needs to resort to N-body

simulations, which start from an initial configuration of dark matter and baryon particles and evolve those components via collisionless gravitational dynamics, hydrodynamics and effective subgrid models. While a proper overview of the simulated physics and numerical optimization schemes are beyond the scope of this thesis, see Dolag et al. (2008); Vogelsberger (2020) for reviews, we note that in recent years it has become increasingly important to include baryonic physics in simulations that are used to calibrate models for cosmological analyses.

### 1.2.4 Density field statistics

As we have seen in §1.2.1 the theory of inflation predicts that the initial conditions originate in a quantum process. Therefore any information we can retrieve from observations will also have to be phrased in a stochastic context. Focusing on the density perturbation $\delta$ we can write down a probability density functional (pdf) $\mathbb{P}\left[\delta(\mathbf{x}_1), \cdots, \delta(\mathbf{x}_n)\right]$ for the overdensity field to be described by the function $\delta$ having values $\delta(\mathbf{x}_i)$ at positions $\mathbf{x}_i$. To find the statistical properties of this field it is useful to define the $n-$point correlators $\xi^{(n)}$ that average the ensemble over products of the field at different spatial positions:

$$\begin{aligned}\xi^{(n)}\left(\mathbf{x}_1, \cdots, \mathbf{x}_n\right) &\equiv \langle \delta(\mathbf{x}_1) \cdots \delta(\mathbf{x}_n) \rangle \\ &\equiv \int \mathrm{d}\delta(\mathbf{x}_1) \cdots \mathrm{d}\delta(\mathbf{x}_n)\, \mathbb{P}\left[\delta(\mathbf{x}_1), \cdots, \delta(\mathbf{x}_n)\right] \delta(\mathbf{x}_1) \cdots \delta(\mathbf{x}_n)\,, \end{aligned} \tag{1.22}$$

where in the first equality we have defined the angle brackets as the ensemble average operation. In the following we will denote the case with $n = 2$ simply as the correlation function $\xi$ and the case with $n = 1$ as the ensemble mean $\langle \delta \rangle$.

Taking a second look at Eq. (1.22) this formulation does not appear to be of any practical use, as today's cosmologist has only one sky to observe and will therefore never be able to probe the ensemble averaging operation. Luckily we can still justify the existence of observational cosmology by imposing a set of assumptions on the pdf. In a first step we extend the cosmological principle to the pdf, requiring it to be invariant under spatial translations and rotations. Additionally, we demand the correlators to converge to zero at least exponentially above some spatial scale. These assumptions form the basis of the ergodic theorem (Weinberg, 2008), which tells us that the only price we have to pay when replacing the ensemble average with a spatial one is to get an additional contribution to the variance. The magnitude of this term, which is known as the cosmic variance, decays inversely proportional to the probed cosmological volume.

A further question we might ask ourselves when looking at Eq. (1.22) is how independent these correlators are. Given that the overdensity field has zero mean, the first three correlators carry complementary information. From the fourth order expressions onward this is not true any-

more[3] and we need to take care of all the additional splitted correlators that are spoiling the independence. Subtraction off those terms one defines the connected correlators $\langle\cdots\rangle_c$ and the associated $N$-point correlation functions (NPCFs):

$$\begin{aligned}\xi_n\left(\boldsymbol{x}_1,\cdots\boldsymbol{x}_n\right) &\equiv \langle\delta(\mathbf{x}_1)\cdots\delta(\mathbf{x}_n)\rangle_c \\ &\equiv \langle\delta(\mathbf{x}_1)\cdots\delta(\mathbf{x}_1)\rangle - \sum_{\pi\in P_n}\prod_{\pi_i\in\pi}\left\langle\delta(\mathbf{x}_{\pi_i(1)})\cdots\delta(\mathbf{x}_{\pi_i(|\pi_i|)})\right\rangle_c \,, \end{aligned} \tag{1.23}$$

where we sum over all possible partitions of $\pi$ the set $\{1,\cdots,n\}$, $\pi_i$ denotes the $i$th element of $\pi$ and $|\pi_i|$ stands for the number of elements withing $\pi_i$[4].

Recalling back from §1.2.2 that the perturbation equations are most easily solved in Fourier space we should also define a harmonic space analog related to the N-point correlation functions. These are called the polyspectra $S_n$ and they are defined via:

$$\langle\delta(\mathbf{k}_1)\cdots\delta(\mathbf{k}_n)\rangle_c \equiv (2\pi)^3\delta^{(D)}\left(\sum_{i=1}^n\mathbf{k}_i\right)S_n\left(\mathbf{k}_1,\cdots,\mathbf{k}_n\right) \,, \tag{1.24}$$

where the due to statistical isotropy the spectra depend only on the directions of the $\boldsymbol{k}$s and the delta function is required by statistical homogeneity. In particular we will denote the lowest nonvanishing orders as the Power spectrum $P$, the Bispectrum $B$, and the Trispectrum $T$. We now show that the polyspectra and the correlation function are simply Fourier pairs of each other:

$$\begin{aligned}\langle\delta(\mathbf{k}_1)\cdots\delta(\mathbf{k}_n)\rangle_c &= \int \mathrm{d}^3x_1\cdots\mathrm{d}^3x_n\mathrm{e}^{-i\sum_{j=1}^n\boldsymbol{k}_j\boldsymbol{x}_j}\,\langle\delta(\mathbf{x}_1)\cdots\delta(\mathbf{x}_n)\rangle_c \\ &\equiv \int \mathrm{d}^3r_1\cdots\mathrm{d}^3r_{n-1}\mathrm{d}^3x_n\mathrm{e}^{-i\sum_{j=1}^{n-1}\boldsymbol{k}_j(\boldsymbol{x}_n+\boldsymbol{r}_j)}\mathrm{e}^{-i\boldsymbol{k}_n\boldsymbol{x}_n} \\ &\qquad\qquad \times\langle\delta(\boldsymbol{x}_n+\boldsymbol{r}_1)\cdots\delta(\boldsymbol{x}_n+\boldsymbol{r}_{n-1})\,\delta(\boldsymbol{x}_n)\rangle_c \\ &= \int \mathrm{d}^3r_1\cdots\mathrm{d}^3r_{n-1}\mathrm{e}^{-i\sum_{j=1}^{n-1}\boldsymbol{k}_j\boldsymbol{r}_j}\,\langle\delta(\boldsymbol{r}_1)\cdots\delta(\boldsymbol{r}_{n-1})\,\delta(\mathbf{0})\rangle_c \\ &\qquad\qquad \times\int \mathrm{d}^3x_n\;\mathrm{e}^{-i\left(\sum_{j=1}^n\boldsymbol{k}_j\right)\boldsymbol{x}_n} \\ &= (2\pi)^3\delta^{(D)}\left(\sum_{i=1}^n\mathbf{k}_i\right)\,\mathcal{F}\left[\xi_n\right]\left(\boldsymbol{k}_1,\cdots,\boldsymbol{k}_n\right) \,, \end{aligned} \tag{1.25}$$

[3] As an example, take $\langle\delta_1\delta_2\delta_3\delta_4\rangle$ and place the point pairs $(1,2)$ and $(3,4)$ infinitely far apart. Assuming ergodicity, this implies that the fourth order correlator cannot be fully independent of the second order one.

[4] To unpack this notation let us explicitly write down $\xi_5$, for which the possible partitions each fall into one of the following structures:

$$(\,i\,|\,j\,k\,l\,m\,)\,,\;(\,i\,j\,|\,k\,l\,m\,)\,,\;(\,i\,|\,j\,|\,k\,l\,m\,)\,,\;(\,i\;j\,|\,k\,l\,|\,m\,)\,,\;(\,i\,|\,j\,|\,k\,|\,l\,m\,)\,,\;(\,i\,|\,j\,|\,k\,|\,l\,|\,m\,)\,.$$

Owing to the fact that the overdensity field has zero mean, only the second partition structure will be nonvanishing. For this structure there are $\binom{5}{2}=10$ different ways on how to arrange the arguments and therefore we find

$$\xi_5\left(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_5\right) = \xi^{(5)}\left(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_5\right) - \left[\xi_2\left(\boldsymbol{x}_1,\boldsymbol{x}_2\right)\xi_3\left(\boldsymbol{x}_3,\boldsymbol{x}_4,\boldsymbol{x}_5\right) + 9\text{ perm.}\right] \,.$$

where in the first equality we re-expressed the overdensity field by using the Fourier convention

$$\delta(\boldsymbol{k}) \equiv \int \mathrm{d}^3x\, \mathrm{e}^{-i\boldsymbol{k}\boldsymbol{x}}\delta(\boldsymbol{x}) \qquad \delta(\boldsymbol{x}) \equiv \int \frac{\mathrm{d}^3x}{(2\pi)^3}\, \mathrm{e}^{i\boldsymbol{k}\boldsymbol{x}}\delta(\boldsymbol{k}) \tag{1.26}$$

and brought the averaging operation inside the integral and in the second one rewrote the correlator in terms of separation vectors from $\boldsymbol{x}_n$. In the third step we made use of statistical homogeneity such that we are able to split the integral. The final step follows when assuming the spectra to be non-vanishing, which implies the delta function. Comparing with Eq. (1.24) we confirm the asserted Fourier correspondence between NPCFs and the polyspectra.

The most simple models for inflation predict the inflaton field to be a Gaussian random field having a scale invariant dimensionless power spectrum[5]. After the end of inflation the inflaton perturbations can directly be related to the gravitational potentials and therefore to $\delta$, making the initial conditions for the density fields Gaussian as well. Using Eq. (1.21) we deduce that for simple inflationary models the matter power spectrum on large modes that enter during matter domination will go as $P(k) \propto k^{n_s}$, while for smaller scales entering during the radiation dominated we have $P(k) \sim k^{-4+n_s}$. In these expressions $n_s$ is known as the spectral index that quantifies the slight departures of the initial power spectrum's scale invariance. Note that as long as the field remains Gaussian all its connected correlators of order $> 2$ vanish and we can extract all its information from the power spectrum (Isserlis, 1918; Wick, 1950). For later evolution stages of the cosmos nonlinear clustering will kick in and induce an increase in the power spectrum on small scales. While the perturbative approaches outlined in §1.2.3 work well up until $k = \mathcal{O}(0.1\ h/\mathrm{Mpc})$ they will break down once the fluid approximation is not valid anymore. If one wants to proceed to nonlinear scales the halo model approach (Neyman & Scott, 1952; Scherrer & Bertschinger, 1991; Cooray & Sheth, 2002) or more flexible fitting functions (Smith et al., 2003; Takahashi et al., 2012; Mead et al., 2015; Smith & Angulo, 2019) can do the job to a sufficient degree. In the left hand side of Fig. 1.4 we compare the linear and nonlinear power spectrum, calculated with CLASS using the Halofit extension of Bird et al. (2012), at three different redshifts. For the linear power spectrum, we explicitly see the power-law limits for large and small scales, as well as the turnover point corresponding to the horizon size at matter-radiation equality. For the nonlinear power, we see that it is the smallest scales that are first to

[5] As the correlation function is dimensionless, the power spectrum will have a dimension of volume. The dimensionless version of the power spectrum $\Delta^2(k) \equiv k^3P(k)/(2\pi^2)$ can then be interpreted as the contribution of power within a logarithmic $k-$interval to the dispersion of overdensity perturbations: $\left\langle \delta(\boldsymbol{x})^2 \right\rangle = \int_{\mathbb{R}_+} \mathrm{d}\ln k\ \Delta^2(k)$. If space expands exponentially during inflation the slow-roll conditions demand the inflaton to roll down the potential with roughly constant 'velocity' such that each decade of scales will have the same power. We therefore get a scale-invariant, constant $\Delta^2$ which is also dubbed a Harrison-Zeldovic power spectrum (Harrison, 1970; Zeldovich, 1972).

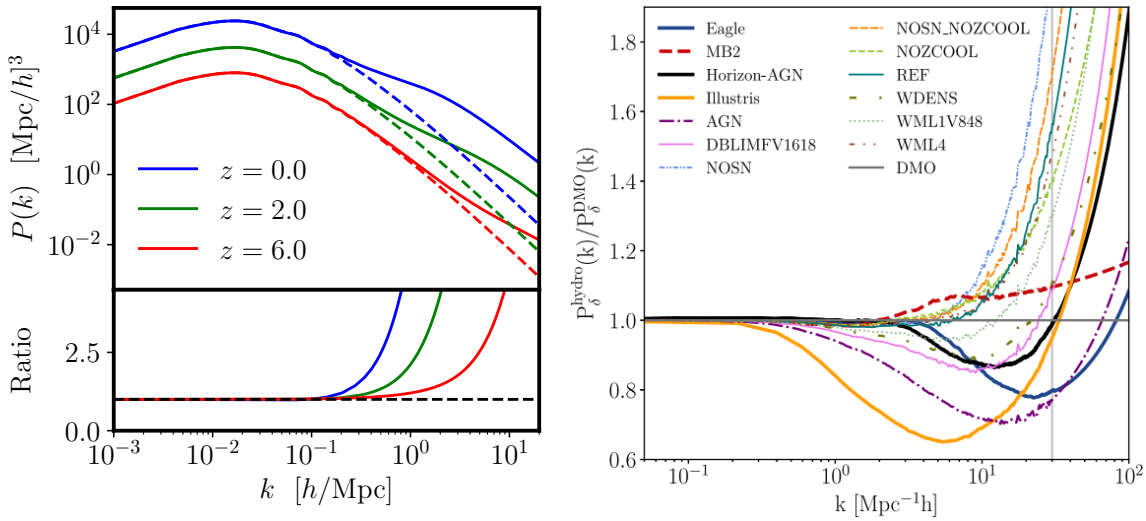


Figure 1.4 *Left hand side*: In the upper panel we show the linear (dashed lines) and nonlinear (solid lines) power spectra for three different redshifts. The lower panel displays the ratio of the nonlinear and linear spectra. *Right hand side*: Impact of baryonic physics on the matter power spectrum. Figure credit: Huang et al. (2019)

gravitationally collapse, in accordance with what one would expect from a hierarchical model of structure formation. Sadly, the story is not over yet as for small scales of $k \gtrsim 1\ h/\mathrm{Mpc}$ one does also need to take into account the impact of baryonic physics, which generally leads to a spoon-like suppression of the power spectrum. In the right hand side plot of Fig. 1.4 we show the ratio of the power spectra from various hydrodynamical simulations and their dark matter only counterparts and we see that there is still a lot of variation between different choices of formulating such models.

## 1.3 Cosmological weak lensing

### 1.3.1 Foundations

The deformation of a light bundle propagating through spacetime can be studied with help of the equation of geodesic deviation (see Fig. 1.5 for a sketch). Let us consider two photons paths intersecting at an angle $\boldsymbol{\theta}$ at $\chi = 0$ and compute the evolution of the comoving separation vector $\boldsymbol{x}$ between the two geodesics. In a simple FLRW background Eq. (1.2) there is no curvature present and the equation can easily be solved as $\boldsymbol{x} \equiv \chi\boldsymbol{\theta}$. On the other hand, if we consider the perturbed metric Eq. (1.14) and assume General Relativity, the perturbation potentials will source

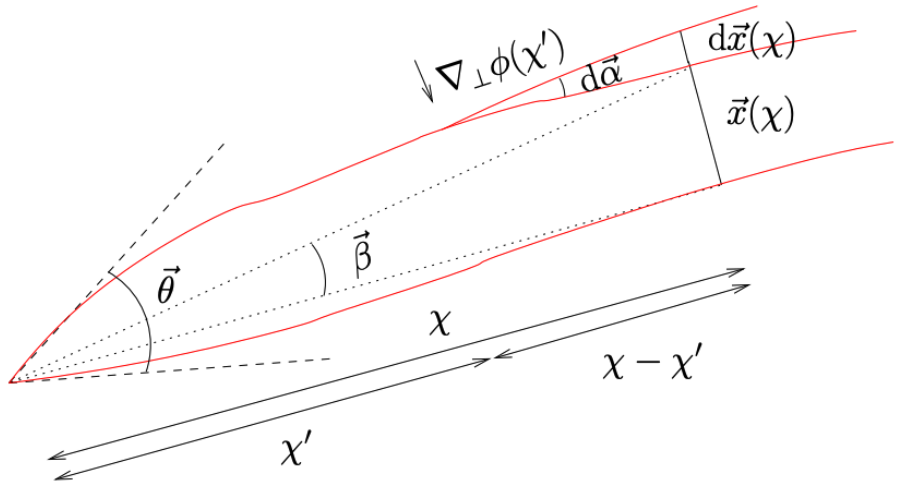


Figure 1.5 Geometry of the lens equation in a cosmological context. In a homogeneous universe the light rays would be seen at comoving separation $\boldsymbol{\beta}$ while the LSS induces a continuous deformation of the path which is observed at an angle $\boldsymbol{\theta}$. Taken from (Kilbinger, 2015).

an inhomogeneity

$$\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}\chi^2} = -\frac{2}{c^2}\left(\nabla_\perp\phi^{(1)}(\boldsymbol{x}(\boldsymbol{\theta},\chi),\chi) - \nabla_\perp\phi^{(0)}(\mathbf{0},\chi)\right) . \tag{1.27}$$

Using the Green's function method and defining the angular separation vector $\boldsymbol{\beta} \equiv \boldsymbol{x}/\chi$ we can rewrite Eq. (1.27) as an integral equation, the lens equation,

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \underbrace{\frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{\chi-\chi'}{\chi'}\left(\nabla_\perp\phi^{(1)}(\boldsymbol{x}(\boldsymbol{\theta},\chi'),\chi') - \nabla_\perp\phi^{(0)}(\mathbf{0},\chi')\right)}_{\equiv\boldsymbol{\alpha}} , \tag{1.28}$$

that quantifies the difference $\boldsymbol{\alpha}$ between the true ($\boldsymbol{\beta}$) and observed position ($\boldsymbol{\theta}$) of a light source in the sky. Expanding the lens equation around $\boldsymbol{\theta}$ to leading order defines the Jacobi matrix

$$\mathcal{A}_{ij}(\boldsymbol{\theta},\chi) \equiv \frac{\partial\beta_i}{\partial\theta_j} \tag{1.29}$$

$$= \delta_{ij} - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{(\chi-\chi')\chi'}{\chi}\phi_{,ik}(\boldsymbol{x}(\boldsymbol{\theta},\chi'),\chi')\mathcal{A}_{kj}(\boldsymbol{\theta},\chi') \tag{1.30}$$

$$\approx \delta_{ij} - \underbrace{\frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{(\chi-\chi')\chi'}{\chi}\phi_{,ij}(\chi'\boldsymbol{\theta},\chi')}_{\equiv\psi(\boldsymbol{\theta},\chi)} , \tag{1.31}$$

where in going to the second line we applied the chain rule and in the third line employed the Born approximation[6] and defined the lensing potential $\psi$.

[6] The Born series is a perturbative solution to the implicit equation Eq. (1.28) in which one expands the photon trajectory $\boldsymbol{x}$ and iteratively uses those coefficients to find the next order solution. As in our case the perturbation variable is the gravitational potential and we consider the weak field limit one should expect one shold expect the higher order terms do decrease quickly. The Born approximation is defined as only taking into account the leading order term in the full series. For the Jacobi matrix this implies $\boldsymbol{x}^{(0)} = \chi\boldsymbol{\theta}$ and results in performing the integration along the unperturebed photon trajectory.

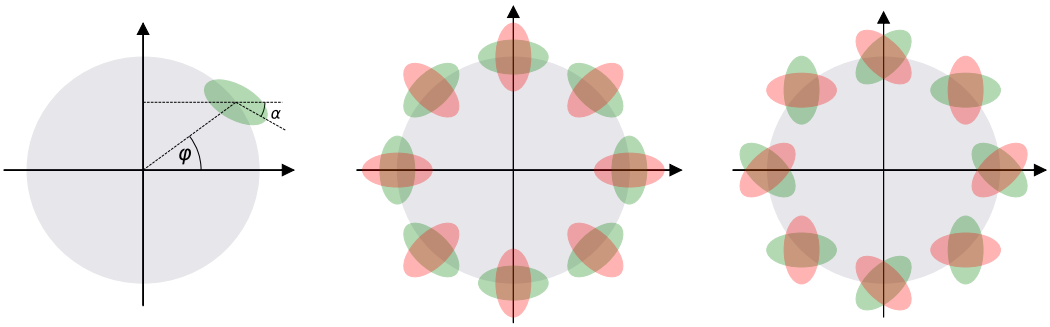


Figure 1.6 *Left hand side:* Definition of the angles $\alpha$ and $\varphi$. In this specific example $\alpha < 0$ and hence $\gamma_t > 0$. *Right hand side:* Definitions of the tangential shear $\gamma_t$ (middle) and cross shear $\gamma_\times$ (right). Green color indicates maximal positive value while red corresponds to maximal negative value.

We can find a useful parametrization of this result when noting that the Jacobi matrix has three independent components such that it can be decomposed in an isotropic and a distortion part

$$\mathcal{A} \equiv (1-\kappa)\,\mathbb{1}_2 + \begin{pmatrix} -\gamma_1 & -\gamma_2 \\ -\gamma_2 & \gamma_1 \end{pmatrix} \quad \rightarrow \quad \lambda_\pm = 1 - \kappa \pm |\boldsymbol{\gamma}|\,, \tag{1.32}$$

where we defined the convergence $\kappa$ and the shear polar $\boldsymbol{\gamma} \equiv \gamma_1 + i\gamma_2 \equiv |\boldsymbol{\gamma}|\mathrm{e}^{2i\varphi}$ and the $\lambda_\pm$ denote the eigenvalues of $\mathcal{A}$. Recalling that Liouville's Theorem guarantees surface brightness conservation for gravitational lensing, the Jacobi matrix $\mathcal{A}$ acts as a linear operator mapping the shapes of images to sources. Inverting this relationship we can show that lensing affects a circular light bundle $\delta\boldsymbol{\beta} \equiv R\binom{\sin\varphi}{\cos\varphi}$ as $\delta\boldsymbol{\theta} = \mathcal{A}^{-1}\delta\boldsymbol{\beta}$ which results in two effects.

- Gravitational lensing maps circular objects into ellipses, such that the ellipticity $\boldsymbol{\epsilon}$ of $\delta\boldsymbol{\theta}$ depends on the reduced shear $\boldsymbol{g} \equiv \frac{\boldsymbol{\gamma}}{1-\kappa}$:

$$\boldsymbol{\epsilon} \equiv \frac{a-b}{a+b} = \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} = \frac{\boldsymbol{\gamma}}{1-\kappa}$$

- Gravitational lensing changes the area $A$ of the source. In the weak limit the magnification $\mu$ is solely induced by the convergence:

$$\mu \equiv \frac{A^{\mathrm{image}}}{A^{\mathrm{source}}} = \frac{ab\pi}{R^2\pi} = \frac{1}{\lambda_+\lambda_-} = \frac{1}{(1-\kappa)^2 - |\boldsymbol{\gamma}|^2} \approx 1 + 2\kappa$$

In order to define shear in a coordinate frame independent fashion we introduce a mapping from the shear components $(\gamma_1, \gamma_2)$ to the so called tangential and cross shear components $(\gamma_t, \gamma_\times)$

in terms of a direction $\varphi$ as

$$\gamma_t(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \equiv -\Re\left[\gamma e^{2i\varphi}\right] = -|\gamma|\cos[2(\alpha-\varphi)] = -(\gamma_1\cos(2\varphi) + \gamma_2\sin(2\varphi)) \tag{1.33}$$

$$\gamma_\times(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \equiv -\Im\left[\gamma e^{2i\varphi}\right] = -|\gamma|\sin[2(\alpha-\varphi)] = -(\gamma_1\sin(2\varphi) - \gamma_2\cos(2\varphi))\,, \tag{1.34}$$

where in the second equality the direction is the angle of the vector $\boldsymbol{\theta} - \boldsymbol{\theta}_0$ and $\alpha \equiv \measuredangle(\boldsymbol{v}_{\lambda_+})$ is the rotation angle induced by the shear. Note that due to the minus sign in the argument of the trigonometric functions this measure indeed is independent of any chosen (orthogonal) coordinate basis. The minus sign in front is convention - i.e. it assigns positive values to the tangential shear when the rotated ellipse is aligns tangentially to the isodensity contours of a positive mass / overdensity distribution.

Let us now look at the convergence in a bit more detail. Comparing Eqs. (1.31) and (1.32) we can relate $\kappa$ to the lensing potential and from there successively find

$$\begin{aligned}\kappa(\boldsymbol{\theta}, \chi) &= \frac{\Delta\psi(\boldsymbol{\theta}, \chi)}{2} \\ &= \frac{1}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{(\chi-\chi')\chi'}{\chi}\Delta^{(3)}\phi(\chi'\boldsymbol{\theta}, \chi') \\ &= \frac{3}{2}\Omega_{\mathrm{m},0}\left(\frac{H_0}{c}\right)^2 \int_0^\chi \mathrm{d}\chi' \frac{(\chi-\chi')\chi'}{\chi a(\chi')}\delta(\chi'\boldsymbol{\theta}, \chi')\,,\end{aligned} \tag{1.35}$$

where in the second step we defined $\Delta^{(3)} \equiv \Delta + \partial_\chi^2$ and assumed that only the derivatives perpendicular to the line of sight contribute to the convergence (White & Hu, 2000); for the final step we substituted the Poisson equation Eq. (1.21). If we observe multiple galaxies following a redshift distribution $n(z)$ we obtain the effective convergence as[7]

$$\kappa(\boldsymbol{\theta}) \equiv \int_0^\infty \mathrm{d}\chi\, n_\chi(\chi)\, \kappa(\boldsymbol{\theta}, \chi) = \int_0^\infty \mathrm{d}\chi\, G(\chi)\delta(\chi\boldsymbol{\theta}, \chi) \tag{1.36}$$

$$G(\chi) \equiv \frac{3}{2}\Omega_{\mathrm{m},0}\left(\frac{H_0}{c}\right)^2 \frac{\chi}{a(\chi)}\int_\chi^\infty \mathrm{d}\chi' n_\chi(\chi')\frac{(\chi'-\chi)}{\chi'}\,, \tag{1.37}$$

where $G$ denotes the effective lensing weight at comoving distance $\chi$. Summarizing, we have found that the integrated large-scale structure along the line of sight induces an isotropic stretching of light bundles. Due to its dependence on $\delta$ we expect $\langle\kappa\rangle \equiv 0$ and need to go to at least second order statistics for retrieving cosmological information.

### 1.3.2 Weak lensing statistics

In complete analogy to the matter spectra, we define the convergence power spectrum $C_{\kappa\kappa}$ as

$$\langle\kappa(\boldsymbol{\ell})\kappa^*(\boldsymbol{\ell}')\rangle \equiv (2\pi)^2\delta^{(D)}(\boldsymbol{\ell}-\boldsymbol{\ell}')C_{\kappa\kappa}(l) \tag{1.38}$$

[7] The double integral only covers the region where $\chi > \chi'$; hence we can first change the order of integration, $\int_0^\infty \mathrm{d}\chi \int_0^\infty \mathrm{d}\chi' \theta^{(H)}(\chi-\chi') = \int_0^\infty \mathrm{d}\chi' \int_{\chi'}^\infty \mathrm{d}\chi$, and afterwards swap $\chi \leftrightarrow \chi'$.

and the higher order spectra in a similar way. In this expression, $\boldsymbol{\ell}$ denotes the two dimensional analog of the wave vector $\boldsymbol{k}$. To relate the convergence power spectrum to the matter power spectrum we make use of two further approximations: First, we take the flat-sky approximation and secondly we assume that the kernel Eq. (1.37) does only vary slightly on the coherence scale of the included LSS. These second assumption forms the basis of the Limber approximation (Limber, 1953; Kaiser, 1992), in which only wavemodes perpendicular to the line of sight contribute to the convergence power spectrum, and we obtain

$$C_{\kappa\kappa}(\ell) = \int_0^\infty \mathrm{d}\chi \frac{G^2(\chi)}{\chi^2} P\left(\frac{\ell}{\chi}, \chi\right) \; ; \tag{1.39}$$

for the computational details we refer to Appendix 1.A. By comparing Eqs. (1.31) and (1.32) we can relate the Fourier space representations of the convergence and shear as $\gamma = ((\ell_1 + i\ell_2)^2/\ell^2)\,\kappa$ and thus we have

$$\langle \gamma(\boldsymbol{\ell})\gamma^*(\boldsymbol{\ell})\rangle = \frac{(\ell_1 + i\ell_2)^2(\ell_1 - i\ell_2)^2}{\ell^4} \langle \kappa(\boldsymbol{\ell})\kappa^*(\boldsymbol{\ell})\rangle = \langle \kappa(\boldsymbol{\ell})\kappa^*(\boldsymbol{\ell})\rangle \quad \Rightarrow \quad C_{\kappa\kappa}(\ell) = C_{\gamma\gamma}(\ell) \; ,$$

meaning that we can infer the theoretical convergence power spectrum from measurements of the power spectrum associated with the measurable shear correlation functions. As the shear is a polar and the convergence a scalar we can construct four correlation functions from the tangential and cross shear components and we expect only the tangential correlator to carry cosmological information. As even in the absence of systematics the other correlators might not vanish due to noise, it is a necessary task to construct measures that only project onto the relevant parts and relate them back to the matter power spectrum. One of those measures is the aperture mass (Schneider et al., 1998) which we will introduce in section §2.3.2.

### 1.3.3 Challenges for Weak Lensing Surveys

Our treatment so far has been very idealized and agnostic towards many theoretical and observational challenges that need to be taken into account when inferring cosmology from the images of galaxies.

On the theoretical side, it is important to make sure that all the employed approximations do just have an effect that is small in comparison to the statistical error budget of the data. In particular, on large scales and for tomographic analysis choices the flat-sky and Limber approximations are not applicable anymore and one needs to take into account correction factors, or make use of the full-sky expressions, which are numerically more demanding (Kilbinger et al., 2017; Fang et al., 2020). On the small scales one needs to take into account corrections that result from dropping the Born and reduced shear approximations (Krause & Hirata, 2010), as well as the

theoretical uncertainties in modelling the nonlinear matter power spectrum itself. Finally, one does also need to take into account the corrections induced by intrinsic alignments, which are a physical effect resulting in nearby galaxies to be correlated and therefore artificially increasing the lensing signal (Catelan & Porciani, 2001; Hirata et al., 2007; Blazek et al., 2019). In the end, a careful analysis of the combined theoretical uncertainty is translated to a range of scales that can be used for constraining cosmological parameters.

On the observational side one needs to perform a number of steps for translating pixelated images to a reliable shape catalog; while a discussion of the this process is beyond the scope of this thesis we refer to Mandelbaum (2018) for a recent review. Finally, it is required to being able to accurately measure the redshift distribution $n(z)$ of the galaxies. As most weak lensing surveys do not measure the full galaxy spectrum, but instead opt for increasing the number of observed galaxies by using a set of broad band filters, the redshifts need to be estimated from just a few flux measurements. There are two main methods aiming to provide accurate estimates of those photo-zs: Template fitting and machine learning.

In the template fitting approaches (Benitez, 2000; Bolzonella et al., 2000; Ilbert et al., 2006) one uses a set of realistic templates to model a galaxies' spectral energy distribution (SED); redshifting the templates one obtains a synthetic photometry prediction for the survey. The redshift of the best fitting SED template does then provide the photometric redshift prediction. One positive feature of this method is that it is physically motivated and does not formally require additional spectroscopy. On the other hand, if the set of template SEDs is not representative of the observed galaxy SEDs, the predictions are much more prone to errors. Machine learning approaches (Collister & Lahav, 2004; Wadadekar, 2005; Carrasco Kind & Brunner, 2014; Sadeh et al., 2016) make use of a representative training set of galaxies from which they learn a global mapping between photometry and redshifts. While not explicitly relying on individual SED templates, the lack of sufficiently many realistic galaxy spectra will hinder the training set to be representative, such that the generalization to the actual galaxies in the survey might induce biases.

For recovering the redshift distribution function $n(z)$ from the individual photometric redshift estimates stacking does not pose a mathematically viable option (Malz, 2021) and further calibration steps become necessary. One possibility (Lima et al., 2008) is given by reweighting a spectroscopic sample such that it matches the observed properties (i.e. colors and magnitudes) of the photometric one. The reweighted $n(z)$ of the spectroscopic sample is then assumed to be an unbiased estimate of the redshift distribution of the photometric sample. An alternative method (Newman, 2008; Menard et al., 2013) makes use of the angular cross-correlations between the photometric sample and any sufficiently overlapping and deep spectroscopic sample. Measuring

this excess clustering in narrow redshift slices then allows to infer the $n(z)$ of the photometric sample. As both these methods are based on multiple nontrivial assumptions, recent surveys have used used both approaches or more sophisticated combinations to calibrate a reliable $n(z)$ and to assess its uncertainties.

## 1.4 Statistical methods

### 1.4.1 Preliminaries

For all empirical sciences, at some point it becomes necessary to check whether our theoretical models are compatible with observed data. The natural and unique formal language to carry out such a judgement is given by probability theory (Cox, 1946; Jaynes, 2003; Van Horn, 2003) and for this work we will choose the Bayesian framework as our tool.

The main probability $\mathbb{P}\left(\mathcal{H} \,\middle|\, \boldsymbol{d}, \mathcal{I}\right)$ we are after governs how likely a hypothesis $\mathcal{H}$ is in the light of some data $\boldsymbol{d}$ and previous information $\mathcal{I}$. Applying Bayes theorem then allows us to rephrase this probability in terms of a forward model from hypothesis to data, a prior and a normalizing factor:

$$\mathbb{P}\left(\mathcal{H} \,\middle|\, \boldsymbol{d}, \mathcal{I}\right) = \frac{\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \mathcal{H}, \mathcal{I}\right) \mathbb{P}\left(\mathcal{H} \,\middle|\, \mathcal{I}\right)}{\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \mathcal{I}\right)} \,. \tag{1.40}$$

Specializing to the discipline of parameter inference we furthermore assume that the hypothesis $\mathcal{H}$ consists of a theoretical model $\mathcal{M}$ which can be described in terms of a set of parameters $\boldsymbol{\theta}$ such that the probability of $\boldsymbol{\theta}$ conditioned on the data, model and background knowledge becomes

$$\mathbb{P}\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{d}, \mathcal{M}, \mathcal{I}\right) = \frac{\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \boldsymbol{\theta}, \mathcal{M}, \mathcal{I}\right) \; \mathbb{P}\left(\boldsymbol{\theta} \,\middle|\, \mathcal{M}, \mathcal{I}\right)}{\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \mathcal{M}, \mathcal{I}\right)} \,. \tag{1.41}$$

In the scientific context, finding a reasonable expressions for the likelihood $\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \boldsymbol{\theta}, \mathcal{M}, \mathcal{I}\right)$ requires both, a theoretical model that is testable by the observations, as well as a sufficient understanding of the measurement process itself. On the other hand, the prior $\mathbb{P}\left(\boldsymbol{\theta} \,\middle|\, \mathcal{M}, \mathcal{I}\right)$ may be set to the posterior of earlier experiments or it can be chosen as an uninformative distribution. For the remainder of this thesis we will neglect the evidence $\mathbb{P}\left(\boldsymbol{d} \,\middle|\, \mathcal{M}, \mathcal{I}\right)$, but only note that it also is an important quantity, i.e. for model comparison problems.

### 1.4.2 Markov chain Monte Carlo methods

Theoretically, once Eq. (1.41) has been written down the problem of parameter inference is solved and the relation just needs to be evaluated for sufficiently many values of $\boldsymbol{\theta}$. Practically, the story is not over yet. For example, let us take a very naive approach and evaluate $\boldsymbol{\theta} \in \mathbb{R}^d$ on a regular grid spanning some volume $\Omega \subset \mathbb{R}^d$ with $n$ points in each direction. Adopting such a brute-force strategy is flawed in two fundamental ways: On the one hand, the number of evaluations scales as $\mathcal{O}(n^d)$ such that high dimensional parameter spaces cannot be probed. Even worse, we do not a priori know whether the probed volume $\Omega$ does actually include any high probability region. Markov Chain Monte Carlo methods comprise of multiple algorithms that deal with aforementioned problems in a more efficient and controlled way.

In a general setup, any MCMC algorithm that aims at sampling from a distribution $\pi$ consists of a stochastic process that generates a homogeneous Markov chain of variables $\boldsymbol{\theta}$. Such a chain consists of a starting position $\boldsymbol{\theta}^{(0)}$ drawn from $\pi$, as well as a transition kernel $T$ being defined as[8]

$$T(\boldsymbol{\theta}^{(i)} \to \boldsymbol{\theta}^{(i+1)}) \equiv \mathbb{P}\left(\boldsymbol{\theta}^{(i+1)} \,\middle|\, \boldsymbol{\theta}^{(i)}, \cdots, \boldsymbol{\theta}^{(0)}\right) = \mathbb{P}\left(\boldsymbol{\theta}^{(i+1)} \,\middle|\, \boldsymbol{\theta}^{(i)}\right) ,$$

where the last equality follows from the Markov property. A useful transition kernel should leave the distribution $\pi$ invariant (stationary) and allow for a traversal through the whole parameter space (ergodicity). When those conditions are fulfilled the distribution of elements within the Markov chain will converge towards $\pi$. In practice one usually invokes the detailed balance condition,

$$\pi(\boldsymbol{\theta}')T(\boldsymbol{\theta}' \to \boldsymbol{\theta}) = \pi(\boldsymbol{\theta})T(\boldsymbol{\theta} \to \boldsymbol{\theta}') , \tag{1.42}$$

which states that the transition kernel has no specified direction of 'time'. We can easily show that detailed balance implies stationarity of $\pi$ with respect to the chain:

$$\int \mathrm{d}\boldsymbol{\theta}'\, \pi(\boldsymbol{\theta}')T(\boldsymbol{\theta}' \to \boldsymbol{\theta}) = \int \mathrm{d}\boldsymbol{\theta}'\, \pi(\boldsymbol{\theta})T(\boldsymbol{\theta} \to \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}) \int \mathrm{d}\boldsymbol{\theta}'\, \mathbb{P}\left(\boldsymbol{\theta}' \,\middle|\, \boldsymbol{\theta}\right) = \pi(\boldsymbol{\theta})$$

and all that is left to do is to devise an explicit form of $T$. The Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) is a simple example achieving that and it usually is also involved in more efficient schemes. It works as follows: starting at some position $\boldsymbol{\theta}^{(i)}$ a new location $\boldsymbol{\theta}^{(i+1)}$ is proposed according to some distribution $g(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})$. One then invokes an accept–reject step $\mathbb{P}_A$ that results in either appending $\boldsymbol{\theta}^{(i)}$ (reject) or $\boldsymbol{\theta}^{(i+1)}$ (accept) to the chain. From this algorithm we can write down the associated transition kernel as

[8] Formally, this is incorrect as in our models we deal with a continuous state space and therefore would need to phrase all arguments in terms of measurable sets. As the main line of reasoning is not affected we will trade the lack of mathematical rigor for a more simple presentation.

$T_{\mathrm{MH}}(\boldsymbol{\theta}^{(i)} \rightarrow \boldsymbol{\theta}^{(i+1)}) = g(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})\mathbb{P}_A(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\theta}^{(i)})$ and use it in conjunction with detailed balance to derive an explicit form of the accept−reject step:

$$\mathbb{P}_A\left(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\theta}^{(i)}\right) = \min\left(1, \frac{\mathbb{P}\left(\boldsymbol{\theta}^{(i+1)}\right)}{\mathbb{P}\left(\boldsymbol{\theta}^{(i)}\right)} \frac{g\left(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i+1)}\right)}{g\left(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)}\right)}\right) . \tag{1.43}$$

For the case of a symmetric proposal distribution $g$ this simply implies that the step is always accepted when it lies in a higher probability region, otherwise the proposal might be rejected and the previous (more probable) sample is repeated. Having theoretically solved the stationarity requirement we still need to find a way to practically test whether a chain has converged to $\pi$. One example of such an empirical assessment is given by the Gelman-Rubin diagnostics (Gelman & Rubin, 1992) in which one runs $M$ randomly initialized chains of length $N$ and then compares the variance $W$ within the chains to the variance $B$ between the chains. In particular, we define the potential scale reduction factor

$$\widehat{\mathrm{PSFR}} \equiv \sqrt{\frac{N-1}{N} + \frac{M+1}{M}\frac{\hat{B}}{\hat{W}}} , \tag{1.44}$$

which in the case of convergence should be close to unity. In practice one usually calls chains with $\widehat{\mathrm{PSFR}} \leq 1.1$ for each parameter to have converged, although this value is not motivated on theoretical grounds and should be chosen to relate to the experimenters required level of convergence.

For very high dimensional or multimodal distributions $\pi$ the MH algorithm will struggle to converge in a computationally feasible timescale and more advanced MCMC approaches, such as affine invariant ensemble and tempering methods (Goodman & Weare, 2010; Swendsen & Wang, 1986), or Gibbs, slice and Hamiltonian sampling algorithms (Turchin, 1971; Geman & Geman, 1984; Neal, 2003; Duane et al., 1987) need to be employed. In the following subsection we will explore the Hamiltonian approach in a bit more detail. Other performant non-MCMC Bayesian methods used in astronomical applications consist of nested sampling (Skilling, 2004) and variational inference (Blei et al., 2017) methods.

### 1.4.3 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal, 2012; Betancourt et al., 2014; Betancourt, 2017) is a sampling strategy that invokes methods from classical Hamiltonian dynamics to construct transition kernels that are able to mostly evade the curse of dimensionality from which a vanilla MH sampler usually suffers from. For this subsection, we will match our equations to the physical convention and denote the set of parameters we want to infer as $\boldsymbol{q}$.

At its core, HMC extends our target distribution $\pi(\boldsymbol{q}) \;:\; \boldsymbol{q} \in \mathbb{R}^n$ to a $2n-$dimensional phase space , $\pi(\boldsymbol{p},\boldsymbol{q}) = \pi(\boldsymbol{p}|\boldsymbol{q})\pi(\boldsymbol{q}) \equiv \mathrm{e}^{-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})}$ , where $\mathcal{H}$ plays the role of the Hamiltonian function and $\boldsymbol{p} \in \mathbb{R}^n$ are a set of auxiliary momentum variables[9]. If we assume the conditional to be independent of the positions, we can retain the target distribution by marginalizing $e^{-\mathcal{H}}$ over the momenta: $\int \mathrm{d}\boldsymbol{p}\, \mathrm{e}^{-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})} = \pi(\boldsymbol{q}) \int \mathrm{d}\boldsymbol{p}\, \pi(\boldsymbol{p}|\boldsymbol{q}) = \pi(\boldsymbol{q}) \int \mathrm{d}\boldsymbol{p}\, \pi(\boldsymbol{p}) = \pi(\boldsymbol{q})$. In this thesis we will make the choice that $\pi(\boldsymbol{p}) = -\ln\left[\mathcal{N}(\boldsymbol{p}|\boldsymbol{0}, M)\right]$ and therefore the Hamiltonian becomes:

$$\mathcal{H}(\boldsymbol{q},\boldsymbol{p}) = -\ln \pi(\boldsymbol{q}) - \frac{1}{2}\boldsymbol{p}^T M^{-1}\boldsymbol{p} + \text{const.} \equiv \psi(\boldsymbol{q}) + K(\boldsymbol{p}) \tag{1.45}$$

For generating a new sample $\boldsymbol{q}'$ from a starting position $\boldsymbol{q}$, HMC proposes the following algorithm:

1. Draw a set of momenta $\boldsymbol{p}$ according to $\pi(\boldsymbol{p})$ and interpret the joint set $(\boldsymbol{q},\boldsymbol{p})$ as a point in phase space.
2. Propose a new set of variables $(\boldsymbol{q}',\boldsymbol{p}')$ by evolving the Hamiltonian system for some pseudo-time $t$ according to the equations of motion

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} = \mathbf{M}^{-1}\boldsymbol{p} \tag{1.46}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} = -\nabla_{\boldsymbol{q}}\psi(\boldsymbol{q})\,. \tag{1.47}$$

3. Perform an accept−reject step on the proposed state according to the rule $\mathbb{P}_A(\boldsymbol{q}',\boldsymbol{q}) \equiv T_A(\mathcal{H}-\mathcal{H}') \equiv \min\left(1, \mathrm{e}^{-\mathcal{H}'+\mathcal{H}}\right)$ and update the chain.

In Fig. 1.7 we give a graphical interpretation of the above algorithm for the example of sampling from a one dimensional unit Gaussian distribution; a proof that this procedure satisfies detailed balance is presented in Appendix 1.B. We see that in HMC the simple random walk transition kernel of MH is replaced by the projection of a Hamiltonian trajectory in phase space to the $\boldsymbol{q}$ basis such that subsequent samples are less correlated and due to the conservation of the Hamiltonian we also expect to achieve very high acceptance rates independent of the dimensionality of the problem. For the computation of the Hamiltonian trajectories we still need to find a discretization scheme that obeys the geometric structure of Hamiltonian mechanics - in particular symplecticity. Fortunately, the theory of geometric integration (see Blanes & Casas (2016) for an overview) allows for many such constructions. The most simple (second order) one is given by the leapfrog integrator, which updates the positions and momenta in the following way:

[9] Note that choosing the parametrization to be $\sim e^{-\mathcal{H}}$ is neccessary, as this form does not only constitute a probability measure, but it also guarantees that the distribution restricted to the level sets of $\mathcal{H}$ are invariant under the Hamiltonian flow restricted to that level set, see i.e. (Betancourt et al., 2014) or any textbook on mathematical statistical mechanics.

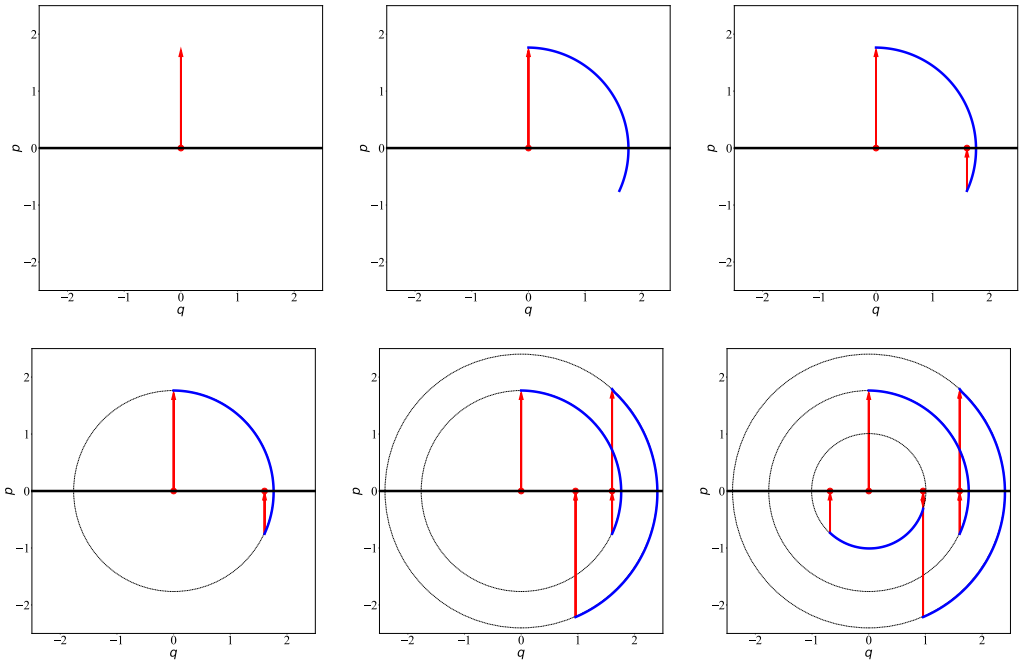

Figure 1.7 Step-by-step walkthrough of an HMC algorithm that samples from a one dimensional unit Gaussian distribution, starting from an initial position $\boldsymbol{q}_0$. In the first step we draw the momentum variable according to the prior distribution to get the initial point $(\boldsymbol{q}_0, \boldsymbol{p})$ in phase space (top left). We then evolve the system according the Hamilton equations of motion (1.46), (1.47) to a new position $(\boldsymbol{q}_1, \boldsymbol{p}')$ in phase space (top middle); in our case $(\boldsymbol{q}, \boldsymbol{p}) \sim \mathcal{G}(0, \mathbb{1}_2)$ and therefore the trajectories will follow a circular path. After checking that energy was conserved along the trajectory we project back onto the $\boldsymbol{q}$-axis to get our second sample $\boldsymbol{q}_1$ (top right). The three panels in the bottom row show the combination of the previous steps for the next two samples $\boldsymbol{q}_2, \boldsymbol{q}_3$ along with their associated energy level sets.

$$
\begin{aligned}
\boldsymbol{p}_{t+\epsilon/2} &= \boldsymbol{p}_t - \frac{\epsilon}{2}\nabla_{\boldsymbol{q}}\psi(\boldsymbol{q}_t) \\
\boldsymbol{q}_{t+\epsilon} &= \boldsymbol{q}_t + \epsilon \mathbf{M}^{-1} p_{t+\epsilon/2} \\
\boldsymbol{p}_{t+\epsilon} &= \boldsymbol{p}_{t+\epsilon/2} - \frac{\epsilon}{2}\nabla_{\boldsymbol{q}}\psi(\boldsymbol{q}_{t+\epsilon}) \,.
\end{aligned}
\tag{1.48}
$$

In these equations there are two parameters which we have not specified yet and which do strongly effect the performance of the sampler. On the one hand, the stepsize $\epsilon$ is responsible for how long the trajectory will be and therefore how correlated subsequent samples are. However, choosing a too large stepsize can result in an inaccurate integration (see left hand side in Fig. 1.8) such that the trajectory leaves the energy hypersurface and therefore such a setting might lead to high rejection rates. On the other hand, the mass matrix $\mathbf{M}$ tells the sampler how

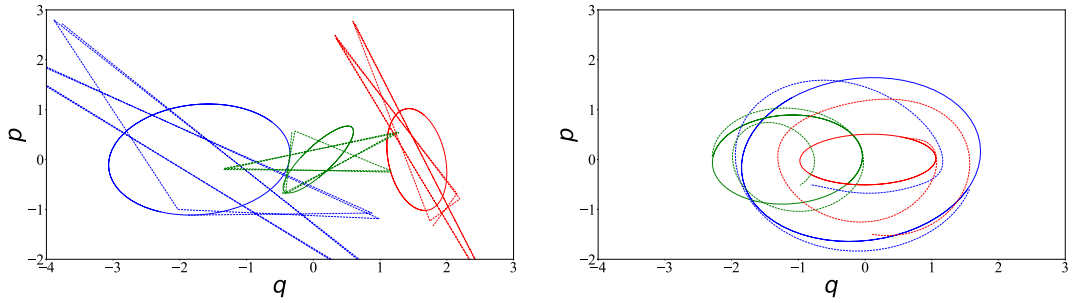


Figure 1.8 Effects of hyperparameters on the Hamiltonian trajectories for the example a 100-dimensional Gaussian with a non-trivial covariance structure. In both plots the elliptic contours of an accurate implementation are shown as the solid lines. For the plot on the left we show a pathological behaviour for three random dimensions when choosing a too large stepsize $\epsilon$ which will result in an inaccurate integration that does not properly trace the energy hypersurface. On the right hand side we show the effect when adopting an inadequate mass matrix for another three random dimensions. Such a difference in the implementation alters the shape of the trajectories and therefore results in a slower convergence of the chains.

different parameters of the chain are 'mixing' during the evolution and one should find a good balance between a mass that allows for fast mathematical operations (in particular inversion, multiplication and Cholesky factorization), but does also approximate the posterior covariance to a sufficient level. A prominent example is the Hessian of the potential, or its diagonal. In the right hand side of Fig. 1.8 we show how the trajectories change when approximating a non-trivial covariance matrix as the identity one.

We note that on the integrator side there have been developed composition methods (Creutz & Gocksch, 1989; Yoshida, 1990) to construct higher order symplectic discretization schemes; while those are more stable they are also numerically much more demanding and it depends on the specific problem on whether they will lead to computational advantage over the standard leapfrog method. On the sampling algorithm itself there exist many additional variations and optimizations, the most heavily employed ones are NUTS (Hoffman & Gelman, 2011), which introduced many hyperparameter autotuning methods, and Riemannian Manifold Hamiltonian Monte Carlo (Girolami & Calderhead, 2011), in which one works within a Hamiltonian system defined on a Riemannian Manifold, therefore taking into account the full posterior curvature information.

## 1.5 Thesis Overview

The research presented in this thesis is organized as follows. In chapter 2 we reproduce Porth et al. (2020). We begin by introducing the the aperture mass statistics and different ways on how to estimate it. We then construct an efficient estimation scheme for the second order statistics and test its accuracy on a large suite mock catalogs constructed from ray-tracing simulations and afterwards apply it to the CFHTLenS data. Chapter 3 has been submitted as Porth & Smith (2021) and presents a more general construction of the estimator introduced in chapter 2 and investigates up to which order one can to expect to measure the aperture mass statistics in a KiDS-1000 like survey. In chapter 4 we assess the information content of projected galaxy fields when employing a forward modelling approach instead of the traditional analysis in terms of the power spectrum.

# Appendix

## 1.A Details on the Limber approximation

In this appendix, which goes along a similar argument as Bartelmann & Schneider (2001), we motivate the expression of the convergence power spectrum within the flat-sky and the Limber approximation. Let us begin by manipulating the expression of the correlator of the convergence (1.36) for a bit:

$$\begin{aligned}
\langle\kappa(\boldsymbol{\theta}_1)\kappa(\boldsymbol{\theta}_2)\rangle &\equiv \int_0^\infty \mathrm{d}\chi_1 \int_0^\infty \mathrm{d}\chi_2\, G(\chi_1)G(\chi_2)\,\langle\delta(\chi_1\boldsymbol{\theta}_1,\chi_1)\delta(\chi_2\boldsymbol{\theta}_2,\chi_2)\rangle \\
&= \int_0^\infty \mathrm{d}\chi_1 \int_0^\infty \mathrm{d}\chi_2\, G(\chi_1)G(\chi_2) \int \frac{\mathrm{d}^3k_1}{(2\pi)^3} \int \frac{\mathrm{d}^3k_2}{(2\pi)^3}\,\langle\delta(k_1,\chi_1)\delta(k_2,\chi_2)\rangle \\
&\qquad \times \mathrm{e}^{i\mathbf{k}_{\perp,1}\cdot\chi_1\boldsymbol{\theta}_1}\mathrm{e}^{ik_{\parallel,1}\chi_1}\mathrm{e}^{i\mathbf{k}_{\perp,2}\cdot\chi_2\boldsymbol{\theta}_2}\mathrm{e}^{ik_{\parallel,2}\chi_2} \\
&= \int_0^\infty \mathrm{d}\chi_1 \int_0^\infty \mathrm{d}\chi_2\, G(\chi_1)G(\chi_2) \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, P(k,\chi_1,\chi_2) \\
&\qquad \times \mathrm{e}^{i\mathbf{k}_\perp\cdot\chi_1\boldsymbol{\theta}_1}\mathrm{e}^{ik_\parallel\chi_1}\mathrm{e}^{-i\mathbf{k}_\perp\cdot\chi_2\boldsymbol{\theta}_2}\mathrm{e}^{-ik_\parallel\chi_2}\,,
\end{aligned} \tag{1.49}$$

where in the going to the second line we decomposed the the wavevector in its perpendicular and parallel components with respect to the line of sight, $\boldsymbol{k} \equiv (\boldsymbol{k}_\parallel, k_\perp)$ and in the last step introduced the unequal time correlator $P(k,\chi_1,\chi_2)$ obeying a Fourier pair relation that is structurally identical to (1.25):

$$\langle\delta(\mathbf{k}_1,\chi_1)\delta(\mathbf{k}_2,\chi_2)\rangle = (2\pi)^3\delta^D(\mathbf{k}_1+\mathbf{k}_2)P(k_1,\chi_1,\chi_2)\,. \tag{1.50}$$

Let us now assume that the correlator vanishes for distances $\theta_{12} \equiv |\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1|$ above some coherence scale $L_{\mathrm{coh.}}$ and also that the light travel time across $2L_{\mathrm{coh.}}$ is sufficiently small such that the kernel function $G$ does not vary significantly across those scales. Defining $\chi \equiv \chi_1$ and $\Delta\chi \equiv \chi_2 - \chi_1 > 0$ we can proceed with the following approximation:

$$\begin{aligned}
\langle\kappa(\boldsymbol{\theta}_1)\kappa(\boldsymbol{\theta}_2)\rangle &\approx \int_0^\infty \mathrm{d}\chi \int_{-L_{\mathrm{coh.}}}^{L_{\mathrm{coh.}}} \mathrm{d}\Delta\chi\, G(\chi)G(\chi+\Delta\chi) \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, P(k,\chi,\chi+\Delta\chi) \\
&\qquad \times \mathrm{e}^{i\mathbf{k}_\perp\cdot\chi\boldsymbol{\theta}_1}\mathrm{e}^{ik_\parallel\chi}\mathrm{e}^{-i\mathbf{k}_\perp\cdot(\chi+\Delta\chi)\boldsymbol{\theta}_2}\mathrm{e}^{-ik_\parallel(\chi+\Delta\chi)} \\
&\approx \int_0^\infty \mathrm{d}\chi\, G^2(\chi) \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, P(k,\chi)\,\mathrm{e}^{i\mathbf{k}_\perp\cdot\chi\boldsymbol{\theta}_1}\mathrm{e}^{ik_\parallel\chi}\mathrm{e}^{-i\mathbf{k}_\perp\cdot\chi\boldsymbol{\theta}_2}\mathrm{e}^{-ik_\parallel\chi}
\end{aligned}$$

$$\times \int_{-L_{\mathrm{coh.}}}^{L_{\mathrm{coh.}}} \mathrm{d}\Delta\chi \, \mathrm{e}^{-ik_\parallel \Delta\chi}$$
$$\approx \int_0^\infty \mathrm{d}\chi \, G^2(\chi) \int \frac{\mathrm{d}^2 k_\perp}{(2\pi)^2} \, P(k_\perp, \chi) \mathrm{e}^{i\mathbf{k}_\perp \cdot \chi(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)} \, , \tag{1.51}$$

where in the approximation we adapted the integration area and in the second one truncated all the Taylor expansions at zeroth order. For the final line we additionally assumed that $k_\parallel L_{\mathrm{coh.}} \gg 1$ such that the innermost integral converges towards a delta function implying that within the Limber approximation no wavemodes parallel to the line of sight contribute to the convergence power spectrum. We now employ again the Fourier pair relation (1.25) to recover the expression (1.39) of the convergence power spectrum in the Limber approximation:

$$\begin{aligned} C_{\kappa\kappa}(\ell) &= \int \mathrm{d}^2\theta_{12} \, \mathrm{e}^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}_{12}} \, \langle \kappa(\boldsymbol{\theta}_1)\kappa(\boldsymbol{\theta}_2)\rangle \\ &\approx \int_0^\infty \mathrm{d}\chi \, G^2(\chi) \int \frac{\mathrm{d}^2 k_\perp}{(2\pi)^2} \, P(k_\perp, \chi) \int \mathrm{d}^2\theta_{12} \, \mathrm{e}^{-i\boldsymbol{\theta}_{12}(\boldsymbol{\ell} - \mathbf{k}_\perp \chi)} \\ &= \int_0^{\chi_{min}} \mathrm{d}\chi \frac{G^2(\chi)}{\chi^2} P\left(\frac{\ell}{\chi}, \chi\right) \, , \end{aligned} \tag{1.52}$$

where in going to the second line we subbed (1.51) and in the final step used the scaling property of the delta function. We note that the corresponding expressions for higher order spectra can be obtained with an identical strategy. For more details on the Limber approximation, its limitations, extensions and applications to tomographic weak lensing surveys we refer to Lemos et al. (2017) and Kilbinger et al. (2017).

## 1.B Proof of detailed balance for HMC

In this appendix we follow the original work of Duane et al. (1987). At first let us find an analytic way to write down the transition kernel associated with the HMC algorithm:

$$T_{\mathrm{HMC}}(\boldsymbol{q} \to \boldsymbol{q}') = \int \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{p}' \, \pi(\boldsymbol{p}) \times T_{\mathrm{Dyn.}} \left[(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{q}', \boldsymbol{p}')\right] \times T_A(\mathcal{H} - \mathcal{H}') \, .$$

Here, $T_{\mathrm{Dyn.}}$ is the transition matrix for the evolution of the dynamic system and $T_A$ is the transition kernel associated with a MH acceptance step. Recalling that $\pi(\boldsymbol{q}) = e^{-\psi(\boldsymbol{q})}$ the detailed balance condition now becomes

$$\begin{aligned} e^{-\psi(\boldsymbol{q})} T_{\mathrm{HMC}}(\boldsymbol{q} \to \boldsymbol{q}') &= e^{-\psi(\boldsymbol{q})} \int \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{p}' \, \pi(\boldsymbol{p}) \, T_{\mathrm{Dyn.}} \left[(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{q}', \boldsymbol{p}')\right] \, T_A(\mathcal{H} - \mathcal{H}') \\ &= \int \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{p}' \, e^{-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})} \, T_{\mathrm{Dyn.}} \left[(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{q}', \boldsymbol{p}')\right] \, T_A(\mathcal{H} - \mathcal{H}') \, . \end{aligned}$$

In theory detailed balance is fulfilled trivially as by the geometric structure of Hamiltonian systems the equations of motion are deterministic and time reversible and the energy is conserved. In the real world this translates to the following set of requirements:

1. The numerical integration of the equations of motions is time reversible - this requires us to use a symplectic discretization scheme like leapfrog:

$$T_{\text{Dyn.}}\left[(\boldsymbol{q},\boldsymbol{p})\to(\boldsymbol{q}',\boldsymbol{p}')\right] = T_{\text{Dyn.}}\left[(\boldsymbol{q}',-\boldsymbol{p}')\to(\boldsymbol{q},-\boldsymbol{p})\right]$$

2. The Hamiltonian is invariant under momentum reversal: $\mathcal{H}(\boldsymbol{q},\boldsymbol{p}) = \mathcal{H}(\boldsymbol{q},-\boldsymbol{p})$

3. The joint MH acceptance $\mathrm{e}^{-\mathcal{H}}T_A$ obeys detailed balance - let us quickly check that:

$$\begin{aligned} e^{-\mathcal{H}}T_A(\mathcal{H}-\mathcal{H}') &\equiv e^{-\mathcal{H}}\,\min\{1,e^{\mathcal{H}-\mathcal{H}'}\} = \min\{e^{-\mathcal{H}},e^{-\mathcal{H}'}\} \\ &= e^{-\mathcal{H}'}\,\min\{e^{\mathcal{H}'-\mathcal{H}},1\} = e^{-\mathcal{H}'}T_A(\mathcal{H}'-\mathcal{H}) \quad \square \end{aligned}$$

4. The integration measure is area preserving: $\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}' = \mathrm{d}(-\boldsymbol{p})\,\mathrm{d}(-\boldsymbol{p}')$ - this is obvious.

Now the proof becomes straightforward:

$$\begin{aligned} e^{-\psi(\boldsymbol{q})}\,T_{\text{HMC}}(\boldsymbol{q}\to\boldsymbol{q}') &= \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\; e^{-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})}\,T_{\text{Dyn.}}\left[(\boldsymbol{q},\boldsymbol{p})\to(\boldsymbol{q}',\boldsymbol{p}')\right]\;T_A(\mathcal{H}-\mathcal{H}') \\ &= \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\; e^{-\mathcal{H}(\boldsymbol{q}',-\boldsymbol{p}')}\,T_{\text{Dyn.}}\left[(\boldsymbol{q}',-\boldsymbol{p}')\to(\boldsymbol{q},-\boldsymbol{p})\right] \\ &\qquad\qquad \times T_A\left[\mathcal{H}(\boldsymbol{q}',-\boldsymbol{p}')-\mathcal{H}(\boldsymbol{q},-\boldsymbol{p})\right] \\ &= \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\; e^{-\mathcal{H}(\boldsymbol{q}',\boldsymbol{p}')}\,T_{\text{Dyn.}}\left[(\boldsymbol{q}',\boldsymbol{p}')\to(\boldsymbol{q},\boldsymbol{p})\right]\;T_A(H'-\mathcal{H}) \\ &= e^{-\psi(\boldsymbol{q}')}\,T_{\text{HMC}}(\boldsymbol{q}'\to\boldsymbol{q}) \quad \square \end{aligned}$$

In going to the second line we used the first three requirements to rewrite the expression and in the next step changed variables as $(-\boldsymbol{p},-\boldsymbol{p}')\to(\boldsymbol{p},\boldsymbol{p}')$ and used the fourth requirement.

# Fast estimation of aperture mass statistics I

# 2

**Abstract**

We explore an alternative method to the usual shear correlation function approach for the estimation of aperture mass statistics in weak lensing survey data. Our approach builds on the direct estimator method. In this paper, to test and validate the methodology, we focus on the aperture mass dispersion. After computing the signal and noise for a weighted set of measured ellipticites we show how the direct estimator can be made into a linear order algorithm that enables a fast and efficient computation. We then investigate the applicability of the direct estimator approach in the presence of a real survey mask with holes and chip gaps. For this we use a large ensemble of full ray-tracing mock simulations. By using various weighting schemes for combining information from different apertures we find that inverse variance weighting the individual aperture estimates with an aperture completeness greater than 70% yields an answer that is in close agreement with the correlation function approach. We then apply this approach to the CFHTLenS as a pilot scheme and find that our method recovers to high accuracy the official result for the variance of both the E and B mode signal. We then explore the cosmological information content of the direct estimator using the Fisher information approach. We show that there is a only modest loss in cosmological information from the rejection of apertures that are of low completeness. This method unlocks the door to efficient methods for recovering higher order aperture mass statistics in linear order operations.

## 2.1 Introduction

Weak gravitational lensing of the light from galaxies is a key tool for constraining the cosmological parameters and distinguishing between competing models of the Universe (Blandford et al., 1991; Kaiser, 1998; Zhang et al., 2007). The first measurements of the correlations in the shapes

of distant background galaxy images are now over two decades old (Bacon et al., 2000; Kaiser et al., 2000; Van Waerbeke et al., 2000; Wittman et al., 2000) and the field of cosmic shear has rapidly matured from these early pioneering studies that mapped of the order a square degree, to the modern surveys KiDS[1], DES[2] and HSC[3], which are mapping thousands of square degrees (Hildebrandt et al., 2017; Troxel et al., 2018; Aihara et al., 2018; Hikage et al., 2019). The next decade will herald in new surveys like *Euclid*[4] and LSST[5] that will map volumes close to the entire physical volume of our observable Universe (Laureijs et al., 2011; LSST, 2009). This will mean that our ability to extract information from such rich data sets will depend almost entirely on our ability to understand and model the complex nonlinear physics involved and our ability to optimally correct or mitigate the systematic errors.

In the last decade, much effort has been invested in extracting cosmological information from the two-point shear correlation functions, and attempts have been made to carefully account for all systematic effects, such as PSF corrections, bias in the ellipticiy estimator, intrinsic alignments (Schneider, 2006b; Massey et al., 2013; Troxel & Ishak, 2015). The two-point shear correlation functions are the lowest order statistics that are of interest and if the convergence field were a Gaussian random field, then they would contain a complete description of the statistical properties of the cosmic shear signal. However, the distribution of observed galaxy ellipticities are non-Gaussian due to various effects: firstly, the nonlinear growth of large-scale structure induces the coupling of density modes on different scales (Bernardeau et al., 2002); secondly, the estimator for shear from ellipticity is a nonlinear mapping (Miralda-Escude, 1991); thirdly, the violation of the Born approximation and the lens-lens coupling also lead to non-Gaussianity in the shear maps (Hilbert et al., 2009). This all leads to a 'flow' of information into the higher order statistics (Taylor & Watts, 2001). A consequence of this is that the errors on measurements of the convergence power spectrum become highly correlated on small scales, limiting the amount of additional information that can be recovered by pushing down to smaller scales (Sato et al., 2011; Hilbert et al., 2012; Kayo et al., 2013; Marian et al., 2013).

The need to go beyond the simple two-point analysis of the data has been highlighted by a number of authors (see for example Sefusatti et al., 2006; Byun et al., 2017). For example, it is well known that $\xi_+$ and $\xi_-$ exhibit a degeneracy between the amplitude of matter fluctuations $\sigma_8$ and the matter density parameter $\Omega_\mathrm{m}$, which scales as $\sigma_8\Omega_\mathrm{m}^{0.5}$. One way to break this degeneracy is by combining the information from the 2-point and 3-point shear correlation functions (Kilbinger

[1] kids.strw.leidenuniv.nl
[2] www.darkenergysurvey.org
[3] hsc.mtk.nao.ac.jp/ssp/
[4] www.cosmos.esa.int/web/euclid
[5] www.lsst.org

& Schneider, 2005; Semboloni et al., 2011; Fu et al., 2014); another way is through adding in the information found in the statistical properties of the peaks in the shear field (Marian et al., 2013; Kacprzak et al., 2016). Given the potential of the non-Gaussian probes to tighten constraints on cosmology and break model and nuisance parameter degeneracies, it is important to study how to optimally measure them, determine how systematics affect them, and to improve the modelling of them. This work will be essential to undertake, if we are to take full advantage of surveys like KiDS, DES, HSC, *Euclid* and LSST.

One of the bottlenecks for accessing the information in the higher-order statistics is that they are challenging quantities to work with. For example, owing to the fact that the shear is a spin-2 field, there are in principle $2^n$ correlation functions to measure for each $n$th order cumulant (Schneider & Lombardi, 2003; Takada & Jain, 2003; Jarvis et al., 2004; Kilbinger & Schneider, 2005). Building the necessary computational tools to measure the 3- and 4-point shear correlation functions is technically challenging and will require large amounts of CPU time to compute all possible configurations (Jarvis et al., 2004; Kilbinger et al., 2014). This is especially true if measurement-noise covariance matrices are to be derived from mock catalogues. In addition, the shear correlation functions are not necessarily the best quantity to measure since they are not E/B mode decomposed (Schneider et al., 2002b; Schneider & Kilbinger, 2007a).

A powerful method to disentangle systematic effects from cosmic shear signals is the E/B decomposition (Crittenden et al., 2001; Schneider et al., 2002b). At leading order, pure weak lensing signals are sourced by a scalar lensing potential, which means that their deflection fields are curl free. Equivalently, the ring-averaged cross component of the shear is expected to be zero (the B mode), while the tangential one contains all the lensing signal (the E mode). Thus B modes enable a robust test for the presence of systematic errors. One method to take advantage of this E/B decomposition is the so-called 'aperture mass statistics' first introduced by Schneider et al. (1998). 'Aperture mass' ($\mathcal{M}_{\rm ap}$) and 'Map-Cross' ($\mathcal{M}_\times$) are obtained by convolving the tangential and cross shear with an isotropic filter function. Therefore by construction they are E/B-decomposed. Taking the second moment leads to the variance of aperture mass, the third to its skewness etc.

The standard approach for measuring the aperture mass statistics in data utilises the fact that, for the flat sky, any $n$-point moment can be expressed in terms of integrals over the $n$-point shear correlation functions, modulo a kernel function (Schneider et al., 2002b; Jarvis et al., 2004). The reason for adopting this strategy stems from the fact that for a real weak lensing survey, the survey mask is a very complicated function: firstly there are survey edges; next, bright stars and their diffraction halo need to be drilled out; chip gaps, if not accounted for in the survey

dither pattern, can lead to additional holes. This small-scale structure in the survey mask means that in order to make the most of the survey data one should measure the correlation functions. However, this approach is not without issue: for example, for the correlation function estimator of the aperture mass dispersion to be accurate and E/B decomposed, one needs to measure $\xi_+$ and $\xi_-$ in angular bins sufficiently fine for the discretisation of the integrals to be reliable (Fu et al., 2014). Further, one also needs to measure the correlation function on scales $\vartheta \in [0, 2\vartheta]$ for the polynomial filter function of Schneider et al. (1998). Owing to galaxy image blending, signal-to-noise issues and the finite size of the survey, the lower bound is never possible and the upper bound means that biases can occur due to edge effects. This leads to so called E/B leakage (Kilbinger & Schneider, 2005). In addition, while the mean estimate is unbiased, the covariance matrix does require one to carefully account for the mask (Schneider et al., 2002a; Friedrich et al., 2016). More recent developments that also make use of the shear correlation functions, while circumventing the issues of E/B leakage on small scales are the ring statistics and COSEBIs (Schneider & Kilbinger, 2007a; Schneider et al., 2010).

In this paper, we take a different approach and explore the direct estimators of the aperture mass statistics, which were first proposed in Schneider et al. (1998). Rather than measuring the correlation functions of the shear polar, only to reduce them by integration to a scalar, we instead directly measure $\mathcal{M}_{\rm ap}$ for a set of apertures and then use an optimised weighting scheme to average the estimates. As we will show in what follows, this approach has some significant advantages over the correlation function approach. In addition to the variance, one can also measure higher order statistics, such as the skewness and kurtosis, with very little additional computational complexity, code modification or CPU expense (see Porth et al. in preparation). These efficiencies will also potentially enable fast computation of covariance matrices and thus rapid exploration of the likelihood surface for such statistics. The possible down sides to this approach, which we explore, are the potential loss of cosmological information arising due to the fact that some incomplete apertures will be rejected. On the other hand, we will also explore the possibility of not rejecting all incomplete apertures, but accepting/weighting apertures based on criteria such as coverage factor and the signal-to-noise. This will lead to E/B leakage, however, as we will show the levels of leakage can be made sufficiently small so that the statistic is accurate within the required errors. As a practical demonstration of this approach we apply it to the CFHTLenS data and present a careful comparison of it with the two-point correlation function method. Lastly, we make use of a large suite of mock catalogues to study the cosmological information content of the two methods for a nominal CFHTLenS like survey and show that there is no substantial loss of information.

The paper breaks down as follows: In §2.2 we define the key theoretical concepts for weak shear and introduce our notation. In §2.3 we define the aperture mass and give expressions for the aperture mass variance in terms of the matter power spectrum, we also give the alternative relation between it and the shear correlation functions. In §2.4 we develop the direct estimator methodology, giving an explicit computation for the mean and variance in the presence of ellipticity weights and also show how the direct estimator can be accelerated and made effectively linear order in the number of galaxies and number of apertures. We discuss various strategies for combining estimates from an ensemble of apertures that give both, high signal-to-noise and a small bias induced by including incomplete apertures. In §2.5 we turn to the analysis of the CFHTLenS data. We give an overview of the data we use and also the mock catalogues that we generate to test for systematic errors. As a preliminary analysis we present the aperture mass maps for the survey. In §2.6 we investigate the bias of the direct estimator induced by the CFHTLenS mask through measuring the aperture mass variance on the mock catalogues and comparing it to the results obtained when using the correlation function method. After determining the weighting scheme that induces the smallest bias we use it to measure the aperture mass variance on the true CFHTLenS data and compare it to the analysis presented in Kilbinger et al. (2013). We also check how the results change when removing blended sources from the data. In §2.7 we use the mock catalogues to investigate the cosmology dependence and the information content of both estimators via the Fisher information. Finally, in §2.8 we summarise our findings, conclude and discuss future work.

## 2.2 Theoretical background

### 2.2.1 Basic cosmic shear concepts

In this paper we are principally concerned with the weak lensing of distant background galaxy shapes by the intervening large-scale structure (Blandford et al., 1991; Kaiser, 1998; Seitz et al., 1994; Jain & Seljak, 1997; Schneider et al., 1998); see Bartelmann & Schneider (2001); Dodelson (2003, 2017) for reviews. The two fundamental quantities describing this mapping from true to observed galaxy images are the convergence $\kappa$ and the shear $\gamma$ which, assuming a metric theory of gravity, are both derived from an underlying scalar lensing potential. In a cosmological setting the convergence at angular position $\boldsymbol{\theta}$ and radial distance $\chi$ can be connected to the density contrast $\delta(\chi\boldsymbol{\theta}, \chi)$ as:

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3}{2}\Omega_{\mathrm{m},0}\left(\frac{H_0}{c}\right)^2 \int_0^\chi \mathrm{d}\chi' \frac{(\chi - \chi')\chi'}{\chi\, a(\chi')}\delta(\chi'\boldsymbol{\theta}, \chi')\,, \tag{2.1}$$

where $\Omega_{\mathrm{m},0}$ is the total matter density, $H_0$ denotes the Hubble constant, $a$ is the scale factor and $c$ the speed of light.

In a real survey we will not necessarily have access to the precise redshifts of each source galaxy. Instead, we will typically have the redshift distribution of sources determined through photometric redshift estimates. Hence, the effective convergence will be obtained by averaging over the source population $p_\chi$:

$$\begin{aligned}\kappa(\boldsymbol{\theta}) &= \int_0^{\chi_{\mathrm{H}}} \mathrm{d}\chi\, p_\chi(\chi)\kappa(\boldsymbol{\theta},\chi) \\ &= \frac{3}{2}\Omega_{\mathrm{m},0}\left(\frac{H_0}{c}\right)^2 \int_0^{\chi_{\mathrm{H}}} \mathrm{d}\chi' \frac{\chi'}{a(\chi')} g(\chi')\delta(\chi'\boldsymbol{\theta},\chi')\,, \end{aligned} \tag{2.2}$$

where $\chi_{\mathrm{H}}$ is the comoving distance to the horizon and the weight function $g(\chi)$ is defined as

$$\begin{aligned} g(\chi') &\equiv \int_{\chi'}^{\chi_{\mathrm{H}}} \mathrm{d}\chi\, p_\chi(\chi)\frac{\chi-\chi'}{\chi} \\ &= \frac{1}{N_{\mathrm{TOT}}}\int_{z(\chi')}^{z_{\mathrm{H}}} \mathrm{d}z\, \frac{\mathrm{d}N(z)}{\mathrm{d}z}\frac{(\chi(z)-\chi')}{\chi(z)}\,, \end{aligned} \tag{2.3}$$

where we used that the weight function can be equivalently written in terms of the differential number counts by noting that $p_\chi(\chi)\mathrm{d}\chi = p_z(z)\mathrm{d}z = (\mathrm{d}N/\mathrm{d}z)/N_{\mathrm{TOT}}$.

In the left panel of Figure 2.1 we show the redshift distribution of galaxies in the CFHTLenS for the four fields W1, W2, W3 and W4 and the total obtained for the combination of all fields. They were obtained by averaging over the BPZ posterior including the lens weights. One can clearly see that there are significant field to field variations in the redshift distributions, with the W2 and W4 fields showing the largest deviations from the mean in the range $z \in [0.2, 0.4]$ for W2 and $z \in [0.5, 0.7]$ for W4, respectively. The shaded region in the plot shows the standard error region on the mean. This was estimated using a jackknife resampling of the data.

The right panel of Figure 2.1 shows the lensing weight function $g(z)$ computed using the estimated distribution function $p(z)$ shown in the left panel for the CFHTLenS. We see that while there are features in the $p(z)$ distribution, these are effectively washed out when computing the lensing weights for the population. In fact, the most significant outlier is the W2 field, which appears to have a slightly high amplitude for redshifts $z > 1.5$.

### 2.2.2 Tangential and cross shear

Owing to the fact that gravity only 'excites' certain shear patterns we wish to rotate the shear into the frame where we can more easily separate out these modes. This is done by decomposing the shear into 'tangential' and 'cross' components. Consider the shear field at a position vector $\boldsymbol{\theta} + \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is an arbitrary location and $\boldsymbol{\theta}$ is a radial vector centred on $\boldsymbol{\theta}_0$. We may rotate

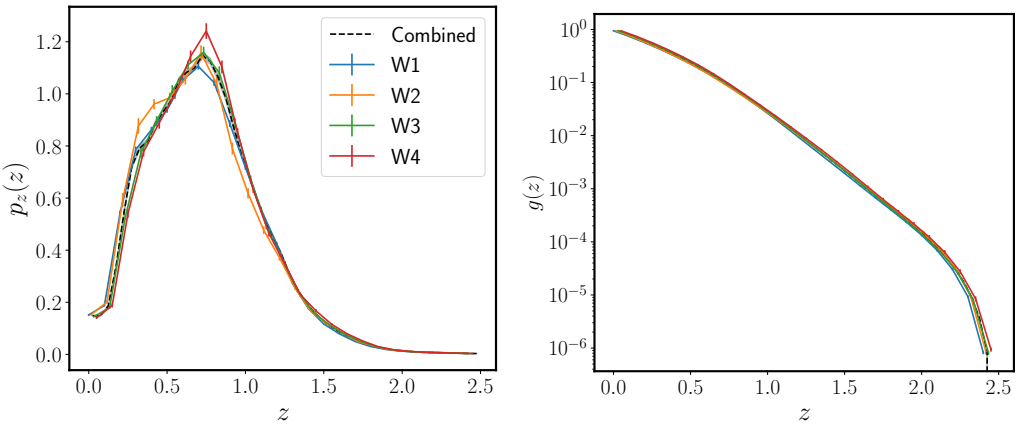


Figure 2.1 *Left hand side:* Redshift probability density distribution of galaxies in the W1, W2, W3 and W4 fields of the CFHTLenS as a function of redshift (i.e. $dN/dz/N_{\rm TOT}$ versus $z$). The coloured solid lines show the results for the individual fields and the dashed black line shows the total for the combined fields. error bars and the grey shaded band show the $1\sigma$ confidence regions, obtained from a jackknife resampling of the survey area. *Right hand side:* The weight function $g(\chi(z))$ given in Eq. (2.3) as a function of redshift for the four CFHTLenSfields and the combined set. Line styles and colours are as in the left panel.

the shear field by the polar angle of the separation vector $\boldsymbol{\theta}$ to obtain the tangential and cross components (Bartelmann & Schneider, 2001):

$$\gamma_{\rm t}(\boldsymbol{\theta};\boldsymbol{\theta}_0) \equiv -\Re\left[\gamma(\boldsymbol{\theta}+\boldsymbol{\theta}_0)\mathrm{e}^{-2i\phi}\right] \ ; \tag{2.4}$$

$$\gamma_{\times}(\boldsymbol{\theta};\boldsymbol{\theta}_0) \equiv -\Im\left[\gamma(\boldsymbol{\theta}+\boldsymbol{\theta}_0)\mathrm{e}^{-2i\phi}\right] \ , \tag{2.5}$$

where $\phi$ is the polar angle associated with the vector $\boldsymbol{\theta}$. The main advantage of this transformation is that for an axially-symmetric mass distribution, the shear is always tangentially aligned relative to the direction towards the origin of the mass distribution and the cross component will vanish. This result is not true for any randomly selected point for the origin. However, if we average the tangential shear over a ring it can be related to the enclosed surface mass density $\overline{\kappa}$: $\langle\gamma_{\rm t}(\boldsymbol{\theta};\boldsymbol{\theta}_0)\rangle_{circ} = \overline{\kappa}(\theta;\boldsymbol{\theta}_0) - \langle\kappa(\theta;\boldsymbol{\theta}_0)\rangle_{circ}$. On the other hand, if we ring average the cross-shear it will vanish: $\langle\gamma_{\times}(\boldsymbol{\theta};\boldsymbol{\theta}_0)\rangle_{circ} = 0$ (Kaiser, 1995; Schneider, 1996).

## 2.3 Aperture mass measures for cosmic shear

In this paper we are primarily concerned with the statistical properties of the ring averaged tangential shear integrated over a filter function with compact support – the aperture mass.

### 2.3.1 Aperture Mass

Aperture mass was developed by Schneider (1996) as a technique for using a weighted set of measured shears within a circular region to estimate the enclosed projected mass overdensity. It can be defined as follows: consider an angular position vector in the survey $\boldsymbol{\theta}_0$, and let us compute the tangential shear field around this point. Aperture mass is now defined as the convolution of the tangential shear with a circularly-symmetric filter function $Q$, with a characteristic scale $\vartheta$, above which the filter functions are typically set to zero. It can be expressed as:

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta}_0;\vartheta) \equiv \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \gamma_{\rm t}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|;\vartheta)\ . \tag{2.6}$$

In a similar vein one can also define the cross component of aperture mass, which we refer to as 'map-cross':

$$\mathcal{M}_{\times}(\boldsymbol{\theta}_0;\vartheta) \equiv \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \gamma_{\times}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|;\vartheta)\ . \tag{2.7}$$

In the absence of systematic errors (B-modes) in the lensing data, map-cross should vanish. $\mathcal{M}_{\rm ap}$ and $\mathcal{M}_{\times}$ are therefore said to be E/B decomposed (Schneider et al., 2002b).

As was proven by Schneider (1996), owing to the fact that the shear and convergence are sourced by the same scalar potential, one can derive an equivalent relation to that above, but computed by convolving the convergence $\kappa$ with a different filter function $U$:

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta}_0;\vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \kappa(\boldsymbol{\theta}_1) U(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|\,;\vartheta)\ . \tag{2.8}$$

It is important to note that the filter functions $Q$ and $U$ are not independent of one another, but are related (Schneider, 1996):

$$Q(\theta;\vartheta) = \frac{2}{\theta^2}\int_0^{\theta} \mathrm{d}\theta'\theta' U(\theta';\vartheta) - U(\theta;\vartheta)\ ; \tag{2.9}$$

$$U(\theta;\vartheta) = 2\int_{\theta}^{\infty} \frac{\mathrm{d}\theta'}{\theta'} Q(\theta';\vartheta) - Q(\theta;\vartheta)\ . \tag{2.10}$$

Also, it is worth noting that the $U$ filter is a compensated function (Bartelmann & Schneider, 2001).

For this work we will be using a polynomial filter function introduced in Schneider et al. (1998):

$$Q(\theta;\vartheta) = \frac{6}{\pi\vartheta^2}\left(\frac{\theta}{\vartheta}\right)^2\left[1-\left(\frac{\theta}{\vartheta}\right)^2\right]\mathcal{H}(\vartheta-\theta)\ , \tag{2.11}$$

where $\mathcal{H}(x)$ is the Heaviside function.

### 2.3.2 Aperture mass variance

For cosmic shear, the expectation of the aperture mass around a randomly selected point vanishes, since $\langle\kappa\rangle = \langle\gamma_{\rm t}\rangle = 0$. Thus, the lowest order non-zero quantity of interest is the variance. Using Eq. (2.6) the variance of the aperture mass can be written as:

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\boldsymbol{\theta}_0;\vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \mathrm{d}^2\boldsymbol{\theta}_2 \, \langle \gamma_{\rm t}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_0)\gamma_{\rm t}(\boldsymbol{\theta}_2;\boldsymbol{\theta}_0)\rangle \\ \times Q(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|;\vartheta)Q(|\boldsymbol{\theta}_2-\boldsymbol{\theta}_0|;\vartheta) \, . \tag{2.12}$$

Using Eq. (2.8) we see that this can be equivalently written as:

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\boldsymbol{\theta}_0;\vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \mathrm{d}^2\boldsymbol{\theta}_2 \, \langle \kappa(\boldsymbol{\theta}_1)\kappa(\boldsymbol{\theta}_2)\rangle \\ \times U(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|\,;\vartheta)U(|\boldsymbol{\theta}_2-\boldsymbol{\theta}_0|\,;\vartheta) \, . \tag{2.13}$$

The Fourier transform of the convergence, $\tilde{\kappa}$, is defined as follows:

$$\kappa(\boldsymbol{\theta}) \equiv \int_{\mathbb{R}^2} \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} \tilde{\kappa}(\boldsymbol{\ell}) \mathrm{e}^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} . \tag{2.14}$$

On using the above transform in Eq. (2.13) we find:

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\boldsymbol{\theta}_0;\vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \mathrm{d}^2\boldsymbol{\theta}_2 \int_{\mathbb{R}^2} \frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2} \frac{\mathrm{d}^2\boldsymbol{\ell}_2}{(2\pi)^2} \, \langle \tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\rangle \\ \times \mathrm{e}^{-i(\boldsymbol{\ell}_1\cdot\boldsymbol{\theta}_1+\boldsymbol{\ell}_2\cdot\boldsymbol{\theta}_2)} U(|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0|\,;\vartheta)U(|\boldsymbol{\theta}_2-\boldsymbol{\theta}_0|\,;\vartheta) \, . \tag{2.15}$$

We next use the statistical homogeneity and isotropy of the correlations of $\kappa(\boldsymbol{\theta})$ to define the convergence power spectrum:

$$\langle \tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\rangle = (2\pi)^2 \delta^{\rm D}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2) C_{\kappa\kappa}(\ell_1) \, . \tag{2.16}$$

On inserting this into Eq. (2.15) and integrating over the Dirac delta function we see that the aperture mass variance can be written:

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\vartheta) = \int_{\mathbb{R}^2} \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} C_{\kappa\kappa}(\ell) \left|\tilde{U}(\boldsymbol{\ell};\vartheta)\right|^2 \, . \tag{2.17}$$

where

$$\tilde{U}(\boldsymbol{\ell};\vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{y} \, \mathrm{e}^{i\boldsymbol{\ell}\cdot\mathbf{y}} U(|\mathbf{y}|\,;\vartheta) \, . \tag{2.18}$$

To progress we need to relate the convergence power to the matter power spectrum $P$. In the small-scale limit and under the Limber approximation, one finds:

$$C_\kappa(\ell) = \frac{9}{4}\Omega_{\rm m,0}^2 \left(\frac{H_0}{c}\right)^4 \int_0^{\chi_{\rm H}} \mathrm{d}\chi \frac{g^2(\chi)}{a^2(\chi)} P\left(\frac{\ell}{\chi},\chi\right) \, . \tag{2.19}$$

In addition, on inserting Eq. (2.11) into Eq. (2.18) one finds that $\tilde{U}(\ell;\vartheta) = 24 J_4(\ell\vartheta)/(\ell\vartheta)^2$. Inserting these results into Eq. (2.17) we have (Schneider et al., 1998):

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\boldsymbol{\theta}_0;\vartheta) = \frac{1}{\vartheta^2}\frac{24^2}{2\pi}\frac{9}{4}\Omega_{\rm m,0}^2 \left(\frac{H_0}{c}\right)^4 \int_0^{\chi_{\rm H}} \mathrm{d}\chi \frac{g^2(\chi)}{a^2(\chi)} \int_0^\infty \mathrm{d}y \frac{J_4^2(y)}{y^3} P\left(\frac{y}{\vartheta\chi},\chi\right) \, .$$

### 2.3.3 Aperture mass variance from shear correlation functions

As discussed earlier, the standard method for estimating the aperture mass variance is through the two-point shear correlation functions. Let us make that connection explicit. The complex shear field has two non-vanishing two-point correlation functions that can be written in terms of its tangential and cross-components as (Schneider et al., 2002b):

$$\xi_+(\theta) \equiv \langle \gamma_{\rm t}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2)\gamma_{\rm t}(\boldsymbol{\theta}_2;\boldsymbol{\theta}_1)\rangle + \langle \gamma_\times(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2)\gamma_\times(\boldsymbol{\theta}_2;\boldsymbol{\theta}_1)\rangle \ ; \tag{2.20}$$

$$\xi_-(\theta) \equiv \langle \gamma_{\rm t}(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2)\gamma_{\rm t}(\boldsymbol{\theta}_2;\boldsymbol{\theta}_1)\rangle - \langle \gamma_\times(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2)\gamma_\times(\boldsymbol{\theta}_2;\boldsymbol{\theta}_1)\rangle \ , \tag{2.21}$$

where in this subsection $\theta \equiv |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$. It can be shown that $\xi_+$ and $\xi_-$ can be written in terms of the convergence power spectrum as:

$$\xi_+(\theta) = \int_{\mathbb{R}^2} \frac{{\rm d}^2\boldsymbol{\ell}}{(2\pi)^2} C_{\kappa\kappa}(\ell) J_0(\ell\theta) \ ; \tag{2.22}$$

$$\xi_-(\theta) = \int_{\mathbb{R}^2} \frac{{\rm d}^2\boldsymbol{\ell}}{(2\pi)^2} C_{\kappa\kappa}(\ell) J_4(\ell\theta) \ . \tag{2.23}$$

Using the orthogonality of the Bessel functions we can invert the above expressions to obtain the convergence power spectrum:

$$C_{\kappa\kappa}(\ell) = 2\pi \int_0^\infty {\rm d}\theta\theta \xi_+(\theta) J_0(\ell\theta) \ ; \tag{2.24}$$

$$= 2\pi \int_0^\infty {\rm d}\theta\theta \xi_-(\theta) J_4(\ell\theta) \ . \tag{2.25}$$

The important consequence of the above relations is that we can now rewrite the aperture mass variance using the shear correlation functions. On substitution of Eqs. (2.24) and (2.25) into Eq. (2.17) one finds (Schneider et al., 2002b):

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\boldsymbol{\theta}_0;\vartheta) = \int_{\mathbb{R}^2} \frac{{\rm d}^2\boldsymbol{\ell}}{(2\pi)^2} \left|\tilde{U}(\boldsymbol{\ell};\vartheta)\right|^2 \pi \left[\int_0^\infty {\rm d}\theta\theta \xi_+(\theta) J_0(\ell\theta) + \int_0^\infty {\rm d}\theta\theta \xi_-(\theta) J_4(\ell\theta)\right] \ . \tag{2.26}$$

On reordering the integrals over $\ell$ and $\theta$, we see that the above can be written more compactly as:

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\boldsymbol{\theta}_0;\vartheta) = \frac{1}{2\vartheta^2} \int_0^\infty {\rm d}\theta\theta \Big[\xi_+(\theta) T_+(\theta|\vartheta) + \xi_-(\theta) T_-(\theta|\vartheta)\Big] \ , \tag{2.27}$$

where

$$T_+(\theta|\vartheta) \equiv \int_0^\infty {\rm d}\ell\ell \left|\tilde{U}(\ell;\vartheta)\right|^2 J_0(\ell\theta) \ , \tag{2.28}$$

$$T_-(\theta|\vartheta) \equiv \int_0^\infty {\rm d}\ell\ell \left|\tilde{U}(\ell;\vartheta)\right|^2 J_4(\ell\theta) \ . \tag{2.29}$$

Once again, on adopting the Schneider polynomial filter Eq. (2.11) we see that the above kernels have an analytic form (Schneider et al., 2002b):

$$T_+(y) = \frac{\mathcal{H}(2-y)}{100\pi}\Bigg[240\left(2-15y^2\right)\cos^{-1}\left(\frac{y}{2}\right) + y\sqrt{4-y^2} \times \left(120 + 2320y^2 - 754y^4 + 132y^6 - 9y^8\right)\Bigg] ; \tag{2.30}$$

$$T_-(y) = \frac{192}{35\pi}y^3\left(1-\frac{y^2}{4}\right)^{7/2}\mathcal{H}(2-y) , \tag{2.31}$$

where in the above $y \equiv \theta/\vartheta$.

There are several important things to note about this: first, for the case of the Schneider polynomial filter function, one needs to measure $\xi_+$ and $\xi_-$ over the range $\theta \in [0, 2\vartheta]$, meaning that we need information from scales close to zero separation. The correlations on small scales can not be accurately measured and will be dominated by image blending issues and shape noise (Kilbinger et al., 2006). Second, the integration to obtain the variance from Eq. (2.27) can only be approximately done using a set of discrete bins which need to be sufficiently dense and non empty. The result of all of this is that there will be some amount of E/B leakage, which will lead to a suppression of the signal on small scales (Kilbinger et al., 2006). The first issue is also a problem for the direct estimator, but the second is not.

## 2.4 Estimating the aperture mass statistics

As discussed in the previous section, there are two approaches to estimating the aperture mass statistics. The correlation function approach outlined in the previous section has been studied in great detail. The direct estimator approach that we explore in this work has not been as well explored, we therefore now describe our extension of this approach in some detail.

### 2.4.1 The direct estimator for the aperture mass dispersion for a single field – including the source weights

Here we follow Schneider et al. (1998), but extend the work to include a set of arbitrary weights for each source galaxy. Let us first introduce the direct estimator of the aperture mass dispersion for a single field. Consider an aperture of angular radius $\vartheta$, centred on the position $\boldsymbol{\theta}_{0,k}$. The aperture contains $N_k$ galaxies with positions $\boldsymbol{\theta}_i$ with complex ellipticities $\epsilon_i$. For the case of weak lensing the observed ellipticities and intrinsic ellipticities $\epsilon_i^{(\mathrm{s})}$ are approximately related though $\epsilon_i = \gamma_i + \epsilon_i^{(\mathrm{s})}$. In complete analogy to the definition of tangential and cross shear defined

in Eqs. (2.4) and (2.5) we define the same quantities for the tangential and cross components of ellipticity: $\epsilon_{\rm t} = -\Re\left[\epsilon\, {\rm e}^{-2i\phi}\right]$ and $\epsilon_\times = -\Im\left[\epsilon\, {\rm e}^{-2i\phi}\right]$, where the polar angle $\phi$ is relative to the origin $\boldsymbol{\theta}_{0,k}$. Our estimator for the aperture mass variance is defined as:

$$\widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})} := (\pi\vartheta^2)^2\, \frac{\sum_i^{N_k}\sum_{j\neq i}^{N_k} w_i\, w_j\, Q_i\, Q_j \epsilon_{{\rm t},i}\, \epsilon_{{\rm t},j}}{\sum_i^{N_k}\sum_{j\neq i}^{N_k} w_i w_j}\,, \tag{2.32}$$

where $w_i$ are weights assigned to the $i$th galaxy, the $Q_i \equiv Q(|\boldsymbol{\theta}_i|;\vartheta)$ and where $\epsilon_{{\rm t},i}$ is the observed tangential ellipticity of the $i$th galaxy measured with respect to the origin $\boldsymbol{\theta}_{0,k}$. Note that since the double sum will occur repeatedly, we will use the short-hand notation $\sum_i^{N_k}\sum_{j\neq i}^{N_k} \to \sum_{i\neq j}$ for brevity. We will also suppress the origin $\boldsymbol{\theta}_{0,k}$ and also take $N_k = N$.

One can show that this provides an unbiased estimator for the true aperture mass dispersion, $\mathbb{E}[$

In Appendix 2.A we also calculate the variance of the estimator Eq. (2.32) and, for a moment suppressing $\vartheta$, we find that it can be written as:

$$\begin{aligned}
{\rm Var}\left[\widehat{M^2_{\rm ap}}\right] = &\frac{\sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l}{\left(\sum_{j\neq i} w_i w_j\right)^2}\left\langle \mathcal{M}^4_{\rm ap}\right\rangle + 4\frac{\sum_{k\neq j\neq i} w_i^2 w_j w_k}{\left(\sum_{j\neq i} w_i w_j\right)^2}\left\langle \mathcal{M}^2_{\rm ap}\mathcal{M}_{\rm s,2}\right\rangle \\
&+ 2\frac{\sum_{j\neq i} w_i^2 w_j^2}{\left(\sum_{j\neq i} w_i w_j\right)^2}\left\langle \mathcal{M}^2_{\rm s,2}\right\rangle + \frac{2\sum_{k\neq i\neq j} w_i w_j^2 w_k}{\left(\sum_{j\neq i} w_i w_j\right)^2}\sigma_\epsilon^2 G\left\langle \mathcal{M}^2_{\rm ap}\right\rangle \\
&+ \frac{2\sum_{j\neq i} w_i^2 w_j^2}{\left(\sum_{j\neq i} w_i w_j\right)^2}\sigma_\epsilon^2 G\left\langle \mathcal{M}_{s,2}\right\rangle + \frac{\sum_{j\neq i} w_i^2 w_j^2}{2\left(\sum_{j\neq i} w_i w_j\right)^2}\sigma_\epsilon^4 G^2 - \left\langle \mathcal{M}^2_{\rm ap}\right\rangle^2,
\end{aligned} \tag{2.33}$$

where $G$ and $\mathcal{M}_{s,2}$ are as defined as:

$$\begin{aligned}
\langle\mathcal{M}_{\rm s,2}\rangle(\vartheta) &\equiv \pi\vartheta^2\int {\rm d}^2\boldsymbol{\theta}\; Q^2(|\boldsymbol{\theta}|;\vartheta)\langle\gamma_{\rm t}^2\rangle(\boldsymbol{\theta})\;; \\
G(\vartheta) &\equiv \pi\vartheta^2\int {\rm d}^2\boldsymbol{\theta}\; Q^2(|\boldsymbol{\theta}|;\vartheta)\;.
\end{aligned} \tag{2.34}$$

Importantly, in the limit where all of the source galaxy weights are equal we recover the expression derived in Schneider et al. (1998).

It is interesting to obtain an approximate form for the above variance. Firstly, let us consider the case where the number of galaxies per aperture is large such that $N \gg 1$, whereupon we see that all of the partial sums are approximately equivalent to the full sums, e.g. $\sum_i\sum_{j\neq i} w_i w_j \approx (\sum_i w_i)^2$. Consequently, all of the prefactors involving the weights can be dramatically simplified to give:

$$\begin{aligned}
\lim_{N\to\infty}{\rm Var}\left[\widehat{M^2_{\rm ap}}\right] \approx &\left\langle \mathcal{M}^4_{\rm ap}\right\rangle + 4S_2\left\langle \mathcal{M}^2_{\rm ap}\mathcal{M}_{\rm s,2}\right\rangle + 2S_2^2\left\langle \mathcal{M}^2_{\rm s,2}\right\rangle \\
&+ 2S_2\sigma_\epsilon^2 G\left\langle \mathcal{M}^2_{\rm ap}\right\rangle + 2S_2^2\sigma_\epsilon^2 G\left\langle \mathcal{M}_{s,2}\right\rangle
\end{aligned}$$

$$+\frac{1}{2}S_2^2\sigma_\epsilon^4 G^2 - \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle^2 \ , \tag{2.35}$$

where we defined $S_2 \equiv \sum_i w_i^2/(\sum_i w_i)^2$. Let us inspect the quantity $S_2$ in more detail: the Cauchy-Schwarz inequality tells us that $\left(\sum_i^N u_i v_i\right)^2 \leq \sum_i^N u_i^2 \sum_j^N v_j^2$, where the elements of the sets $\{u_i\}$ and $\{v_i\}$ are drawn from the reals. If we take $v_i = v$ for any $i$, then we see that $\left(\sum_i^N u_i\right)^2 \leq \sum_i^N u_i^2 N$, which in turn implies $\langle u\rangle^2 \leq \langle u^2\rangle$. On applying this to our ratio $S_2$ we see that:

$$S_2 \equiv \frac{1}{N}\frac{\overline{w^2}}{\overline{w}^2} \geq \frac{1}{N} \tag{2.36}$$

where $\overline{w^2} = \sum_i w_i^2/N$ and $\overline{w} = \sum_i w_i/N$. This insight leads us to make our next approximation, since $\mathcal{M}_{s,2} \sim \mathcal{M}_{\rm ap}^2$ we see that $S_2\mathcal{M}_{s,2} \ll \mathcal{M}_{\rm ap}^2$, since $S_2 \propto 1/N$. This, then, leads us to write:

$$\lim_{N\to\infty} \mathrm{Var}\left[\widehat{M_{\rm ap}^2}\right] \approx \left\langle \mathcal{M}_{\rm ap}^4 \right\rangle + 2S_2\sigma_\epsilon^2 G\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle + \frac{1}{2}S_2^2\sigma_\epsilon^4 G^2 - \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle^2 \ . \tag{2.37}$$

Thirdly, let us further assume that the underlying shear field is Gaussian and hence $\left\langle \mathcal{M}_{\rm ap}^4 \right\rangle = 3\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle^2$. Under these circumstances, which will be fulfilled for large apertures, the variance can be written as:

$$\begin{aligned}\mathrm{Var}\left[\widehat{M_{\rm ap}^2}\right] &\approx 2\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle^2 + 2S_2\sigma_\epsilon^2 G\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle + \frac{1}{2}S_2^2\sigma_\epsilon^4 G^2 \\ &= \left(\sqrt{2}\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle + \frac{1}{\sqrt{2}}S_2 G\sigma_\epsilon^2\right)^2 \ .\end{aligned} \tag{2.38}$$

The first term in the bracket is cosmic variance and the last term denotes the shape noise contribution.

The left panel of Figure 2.2 shows the error on a given estimate from a single aperture, using Eq. (2.38). The right panel shows the corresponding prediction of the signal-to-noise on the aperture mass variance, per aperture, again using Eq. (2.38). In order to generate these prediction we have used (2.20) as a model for the cosmic variance contribution.

### 2.4.2 Acceleration of the direct estimator

If we were to naively implement the direct estimator approach as given by Eq. (2.32) then we see that in order to compute the estimate of the variance for a single aperture we need to compute the sum from $N(N-1)$ galaxies. Thus one might conclude that the method scales as typical $N^2$ pair counting approach for galaxies inside the aperture. However, we now show that the method can be made to scale linearly with the number of galaxies. Let us consider again the estimator from Eq. (2.32), and we notice that if we put back the term that has $i \neq j$ and explicitly subtract

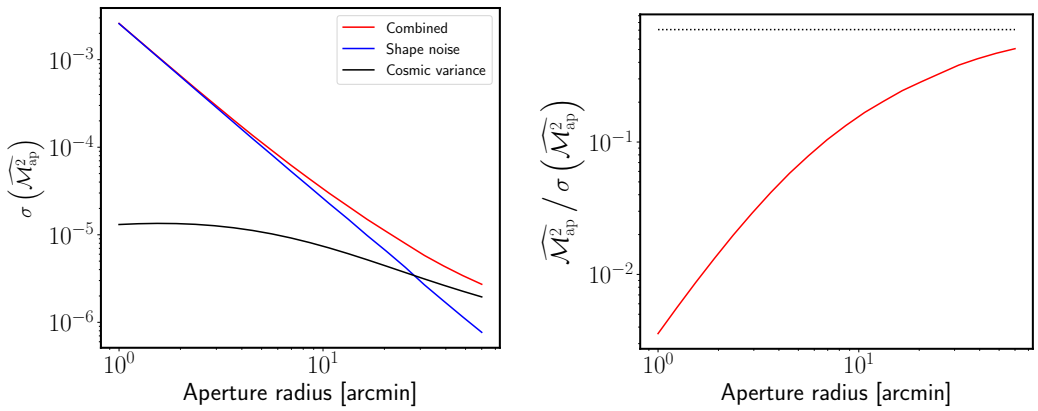


Figure 2.2 *Left hand side:* Contributions to the variance of the estimate of the aperture mass dispersion in the CFHTLenS data for a single aperture. We show the two contributions to the variance as predicted from (2.38) as well as the combined result. Note that in adopting (2.38) we have assumed that the convergence is Gaussianly distributed. *Right hand side:* The theoretical signal-to-noise per aperture as a function of the aperture size $\vartheta$ in the CFHTLenS data. The black dotted line shows the cosmic variance limit.

it then we have:

$$\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})} = (\pi\vartheta^2)^2 \frac{\sum_{i,j} w_i\, w_j\, Q_i\, Q_j\, \epsilon_{{\rm t},i}\epsilon_{{\rm t},j}}{\sum_{i\neq j} w_i w_j} - (\pi\vartheta^2)^2\, \frac{\sum_i w_i^2\, Q_i^2 \epsilon_{t,i}^2}{\sum_{i\neq j} w_i w_j}\,. \tag{2.39}$$

If we now introduce the estimators for aperture mass and $\mathcal{M}_{s,2}$ as discretised versions of their definition,

$$\widehat{M_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})} \equiv (\pi\vartheta^2) \frac{\sum_i w_i\, Q_i \epsilon_{{\rm t},i}}{\sum_i w_i} \tag{2.40}$$

$$\widehat{M_{\rm s,2}(\vartheta|\boldsymbol{\theta}_{0,k})} \equiv (\pi\vartheta^2) \frac{\sum_i w_i^2\, Q_i^2 \epsilon_{{\rm t},i}^2}{\sum_i w_i^2}\,, \tag{2.41}$$

we see that Eq. (2.39) can be rewritten as:

$$\begin{aligned}\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})} &= \frac{1}{1-S_2}\left(\frac{\pi\vartheta^2 \sum_i w_i\, Q_i \epsilon_{{\rm t},i}}{\sum_i w_i}\right)^2 - \frac{\pi\vartheta^2}{\frac{1}{S_2}-1}\,\pi\vartheta^2\,\frac{\sum_i w_i^2\, Q_i^2 \epsilon_{{\rm t},i}^2}{\sum_i w_i^2} \\ &= \frac{1}{1-S_2}\left[(\widehat{M_{\rm ap}})_k^2 - S_2 \pi\vartheta^2 (\widehat{M_{\rm s,2}})_k\right]\,,\end{aligned} \tag{2.42}$$

where for brevity we used the notation $(\widehat{M_{\rm ap}})_k \equiv \widehat{M_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}$ and $(\widehat{M_{\rm s,2}})_k \equiv \widehat{M_{\rm s,2}(\vartheta|\boldsymbol{\theta}_{0,k})}$. Note that both terms in the brackets receive an identical contribution of shape noise, hence the second term should not be neglected.

In general, the estimator Eq. (2.42) is mathematically identical to that of Eq. (2.32) and therefore is also an unbiased estimator for the variance of the aperture mass. However, algorithmically it has a significant advantage in that it is linear in the number of galaxies. This owes to the fact

that all of the terms on the right-hand-side of Eq. (2.42) are linear in $N$. For example, the estimator $(\widehat{M_{\rm ap}})_k$ is linear, so too are the correction factors $(\widehat{M_{\rm s,2}})_k$ and $S$. As we will show in the second paper in this series (Porth et al., in prep.), it can be shown that this can be naturally extended to higher order aperture mass statistics. This acceleration of the method to linear order opens the door to a significant advantage in speed for estimation of aperture mass statistics at all orders.

### 2.4.3 Extending the estimate to an ensemble of fields

The estimator Eq. (2.42) is for a single aperture $k$ and as such will provide a single low-signal-to-noise, albeit unbiased, estimate. We now wish to make use of the full area of the survey available to us. We are therefore confronted as to how to best achieve this. As proposed by Schneider et al. (1998), one simple approach would be to sample well separated apertures such that the shear in one field is statistically independent from another. This would yield the estimator:

$$\widehat{\mathcal{M}^2_{\rm ap}(\vartheta)} = \frac{\sum_k \mathcal{W}_k \widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}}{\sum_k \mathcal{W}_k} \,, \tag{2.43}$$

where $\mathcal{W}_k$ are weights and the sum extends over the $N_{\rm ap}$ apertures. Since the estimates can be considered to be statistically independent then the noise can be minimised by choosing the weights to be given by:

$$\mathcal{W}_k \propto \frac{1}{{\rm Var}\left[\widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}\right]} \,. \tag{2.44}$$

However, this approach would be suboptimal in that it does not take advantage of the full area of the survey. In this case, the signal-to-noise on the estimate for the full field can be achieved by multiplying the estimates for the aperture mass variance per single aperture by the square root of the number of independent apertures.

A much better approach, which takes advantage of a larger portion of the available survey area, is to oversample the apertures, i.e. by placing them on a regular grid with a pixel width of less than twice the aperture radius. Since the estimate for the survey is still given by Eq. (2.43) and since it is a linear combination of the estimates for the single field, it too is unbiased:

$$\begin{aligned}\mathbb{E}\left[\widehat{\mathcal{M}^2_{\rm ap}(\vartheta)}\right] &= \frac{\sum_k \mathcal{W}_k \, \mathbb{E}\left[\widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}\right]}{\sum_k \mathcal{W}_k} \\ &= \frac{\sum_k \mathcal{W}_k \left\langle \mathcal{M}^2_{\rm ap}(\vartheta)\right\rangle}{\sum_k \mathcal{W}_k} = \left\langle \mathcal{M}^2_{\rm ap}(\vartheta)\right\rangle \,. \end{aligned} \tag{2.45}$$

However, the variance of the estimate for the survey does now receive an additional contribution arising from the spatial cross correlation coefficient between apertures and therefore is no longer

trivial to determine. This in turn means that the weights $\mathcal{W}_k$ from Eq. (2.44) are no longer optimal. Computing the optimal weights will be further complicated if we include incomplete apertures in the estimate – which we discuss next[6].

### 2.4.4 Allowing incomplete aperture coverage to increase estimator efficiency

We next turn to the problem of aperture completeness. In real surveys there are regions of the survey that are masked out due to bright stars, chip gaps and the survey boundaries. The question now arises: *what do we do if an aperture has some fraction of its area overlapping with the mask?* The simple answer would be that we exclude all such apertures from the estimator. The problem with this approach is that depending on the size of the aperture this may significantly impact the available survey area from which to compute the estimate and thus make the approach sub-optimal. Here we will explore the idea of effectively including all apertures that fit within the survey boundary, but apply weights to each of the form:

$$\mathcal{W}_k = f(c_k, \mathrm{Var}, \dots) \,, \tag{2.46}$$

where $c_k \equiv A_k/A$ is the completeness factor for the $k$th aperture, where $A_k$ is the available area of the aperture and $A$ is the unmasked area of the aperture, such that we have $c_k \leq 1$. Var is related to the variance of the estimate in the aperture. The ellipsis denotes that in general the weights could depend on other factors. In this work we will explore three distinct choices:

$$\mathcal{W}_k^{(1)} = \mathcal{H}\left(c_k - \alpha\right); \tag{2.47}$$

$$\mathcal{W}_k^{(2)} = \mathcal{H}\left(c_k - 0\right) \mathrm{Var}\left[\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{-1}; \tag{2.48}$$

$$\mathcal{W}_k^{(3)} = \mathcal{H}\left(c_k - \alpha\right) \mathrm{Var}\left[\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{-1}, \tag{2.49}$$

The first case corresponds to accepting all apertures whose completeness factor $c_k \geq \alpha$ and combining them in a simple average with equal weights to arrive at our estimate for $\widehat{M_{\rm ap}^2}$. The second case corresponds to accepting all apertures, irrespective of the completeness factor, but combining all of the estimates using an inverse variance weighted estimate, where the variance is approximated by Eq. (2.38). The third case is simply the product of the first and second cases.

It is important to note that unless $\alpha = 1$ our estimator given by Eq. (2.43) will formally become biased. This means that we will expect some leakage of E/B modes. Postponing a thorough analytical and numerical analysis of incomplete aperture coverage and de-biasing strategies to a companion paper, we will content our selves by investigating the degree of bias that is introduced

[6] Deriving the optimal weighting scheme and aperture mass filter function, that maximizes the signal-to-noise is a topic of great interest, but we shall leave this for a future investigation.

by computing the aperture-cross statistics. For reference, these are defined in direct analogy with Eq. (2.32):

$$\widehat{M_\times^2(\vartheta|\boldsymbol{\theta}_{0,k})} := (\pi\vartheta^2)^2 \, \frac{\sum_{i\neq j} w_i \, w_j \, Q_i \, Q_j \, \epsilon_{\times,i} \, \epsilon_{\times,j}}{\sum_{i\neq j} w_i w_j} \, . \tag{2.50}$$

It can be proven that the expectation of this estimator vanishes; that is provided we have no bias in the estimate we have $\mathbb{E}\left[\widehat{M_\times^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right] = 0$. However, the variance of the estimator does not vanish and it should be given by the pure shape noise contribution to Eq. (2.33). Under the approximations of Eq. (2.38) this is: $\mathrm{Var}\left[\widehat{M_\times^2(\vartheta)}\right] \approx 2S^2\sigma_\epsilon^4 G^2$ per aperture.

We note that it is important to appreciate that the weights $\mathcal{W}_k$ apply to how different fields are combined, and that the weights $w_i$ from Eq. (2.32) apply to how the source galaxies are combined in arriving at an estimate for a single field. We assume that these latter weights have been pre-computed by the method for estimating galaxy ellipticities.

Finally, before moving on, we note that both, the signal-to-noise properties as well as the bias induced by incomplete aperture coverage are likely to depend on the choice of the filter function $Q$ employed in the estimator. Our choice of the polynomial filter (2.11) is mainly motivated by the fact that it only has support within the aperture and therefore induces less bias in small apertures than more spatially extended filters, i.e. the Gaussian filter introduced in Crittenden et al. (2001). We postpone a thorough discussion on how to construct a filter function that has both, high signal-to-noise and low bias to future work.

### 2.4.5 Estimating computational complexity for evaluation of the direct estimator for $\left\langle \mathcal{M}_{\mathrm{ap}}^2 \right\rangle$

Before moving on to the computation of the estimator with real data let us estimate the computational cost for an evaluation of $\widehat{\mathcal{M}_{\mathrm{ap}}^2}$. As described above, the actual implementation is built from a series of algorithmic blocks.

(i) We first construct a KD-tree data structure for the galaxy catalogue.

(ii) The full survey is tiled with overlapping apertures, where the centres are separated by a distance $d$.

(iii) The aperture coverage map is computed to give the $c_k$ values for every aperture.

(iv) For apertures that pass the selection cut ($\mathcal{H}(c_k - \alpha)$), a KD-tree range search locates all galaxies that lie inside the aperture radius $\vartheta$.

(v) Estimate the aperture mass statistics and its variance according to Eq. (2.33) for the $k$th aperture.

(vi) Combine the $N_{\rm ap}$ estimates through a weighted mean of the resulting estimates.

In the above algorithm we shall assume that Steps (i)–(iii), and Step (vi) are performed once and therefore are not the limiting factors for the execution of the code. We do note, however, that the construction of the KD-tree may have a large memory footprint and will take some non-negligible time for the initial construction. The parts of the method that require some consideration are Steps (iv) and (v).

Step (iv) is a range search routine and Step (v) is a routine that evaluates the sums in Eq. (2.42). To compute the complexity for these steps we first identify some parameters: let $p$ specify the order of the statistics; $\vartheta$ describe the aperture radius; and $\zeta$ be a parameter that determines the spacing $d$ between apertures: $d \equiv \vartheta/\zeta$. We note that for a non-overlapping field of apertures whose circumferences just touch each other, we would set $\zeta = 1/2$. Further, the number of apertures is thus a function of $\vartheta$ and $\zeta$, $N_{\rm ap}(\vartheta, \zeta)$. The order for the complexity can thus be computed as follows:

$$\begin{aligned}\mathcal{O}(\widehat{\mathcal{M}^p_{\rm ap}}|p,\vartheta,\zeta) = \sum_{k=1}^{N_{\rm ap}(\vartheta,\zeta)} \Big[&\mathcal{O}(\text{range search}|N_k, N_{\rm nodes}, \vartheta) \\ &+ \mathcal{O}(\text{compute statistic}|N_k, p)\Big] \\ \approx N_{\rm ap}(\vartheta,\zeta)\Big[&\mathcal{O}(\text{range search}|\overline{N}, N_{\rm Tot}, \vartheta) \\ &+ \overline{N}\mathcal{O}(\text{compute statistic}|N=1, p)\Big]\,. \end{aligned} \tag{2.51}$$

The first thing to notice is that the number of apertures scales all of the computations, so if we fix the parameter $\zeta$, then the total number of apertures will scale as $N_{\rm ap} \propto \Omega_{\rm s}/\vartheta^2$, where $\Omega_{\rm s}$ is the survey area. The first term in the square brackets gives the computational time for a range search to deliver back the $N_k$ galaxies in the aperture. If the distribution of source galaxies is roughly randomly distributed on the sky, then we make the approximation $N_k \approx \overline{N} = N_{\rm Tot} A/\Omega_{\rm s}$. Each such range search operation then takes of the order $\mathcal{O}(\log N_{\rm Tot})$ time to execute, but this factor will also scale with the aperture radius and how clustered the source galaxy data are, and also the depth we need to go in the tree walk.

Considering the second term in square brackets, this is the required time for computation of the estimate for the $p$th order aperture mass statistic. As was described earlier for $\widehat{M^2_{\rm ap}}$, the estimator scales linearly with the number of galaxies in the aperture, thus in the second line

Table 2.1 Overview of the CFHTLenS data.

| Field | # of galaxies | Angular Area $\Omega_s$ $[deg]^2$ | $\bar{n}$ $[arcmin]^{-2}$ |
|---|---|---|---|
| W1 | 1871709 | 42.36 | 12.27 |
| W2 | 499372 | 11.72 | 11.84 |
| W3 | 1192084 | 25.23 | 13.12 |
| W4 | 558515 | 12.55 | 12.36 |

we simply scale up the complexity to estimate the statistic for a single galaxy by the number of galaxies in the aperture. As we will explore in our companion paper, owing to this linear scaling, there is no additional overhead in extending the method to compute higher order statistics, beyond the variance, such as the skewness $p = 3$ and kurtosis $p = 4$.

## 2.5 Application to the CFHTLenS data

We now turn to estimating the aperture mass variance in the CFHTLenS data and in a large series of mocks generated by ray-tracing through $N$-body simulations.

### 2.5.1 CFHTLenS shear data

The Canada-France-Hawaii Telescope Lensing Survey (hereafter CFHTLenS) is a weak lensing survey that was completed around 2010. It covers 154 $deg^2$ of the sky in five optical bands $\{u^*,\ g',\ r',\ i',\ z'\}$ with a $\sim 5\sigma$ point source limiting magnitude in the $i'$ band of $i'_{\rm AB} \sim 25.5$. The survey measures galaxy ellipticities for use in weak lensing analysis from multicolour data obtained as part of the CFHT Legacy Survey[7]. The survey data is distributed into four well spaced fields, three of which (W1, W2 & W4) lie close to the equatorial plane, and the third (W3) lies at high declination. Full details of the survey can be found in Heymans et al. (2012).

In this work we make use of the final public data release. The combined data set of W1, W2, W3 and W4 contains ellipticity measurements for 4,121,680 galaxies. In Table 2.1 we provide a summary overview of the data. Associated with each galaxy are: the angular positions RA and DEC, in radians; the $x$- and $y$-pixel coordinates in the projected tangent map; the ellipticity estimates $\epsilon_1$ and $\epsilon_2$, the lens weights $w$; the shear bias correction $c$ and the multiplicative bias correction $m$; photometric redshift estimate $z_{phot}$. Figure 2.1 shows the redshift distribution and lensing efficiency for the sources in each of the four CFHTLenS fields.

[7] www.cfht.hawaii.edu/Science/CFHLS/

Table 2.2 zHORIZON cosmological parameters. Columns are (from left to right): density parameters for matter, dark energy and baryons; the amplitude of the power spectrum; the dark energy equation of state parameters; the spectral index of the primordial power spectrum; the Hubble parameter.

| Model | $\Omega_{\mathrm{m},0}$ | $\Omega_{DE,0}$ | $\Omega_{\mathrm{b},0}$ | $\sigma_8$ | $w_0$ | $w_a$ | $n_s$ | $h$ |
|---|---|---|---|---|---|---|---|---|
| Fiducial | 0.25 | 0.75 | 0.04 | 0.8 | -1 | 0 | 1 | 0.7 |
| $\Omega^{\pm}_{\mathrm{m},0}$ | 0.2/0.3 | 0.8/0.7 | 0.04 | 0.8 | -1 | 0 | 1 | 0.7 |
| $\Omega^{\pm}_{\mathrm{b},0}$ | 0.25 | 0.75 | 0.035/0.45 | 0.8 | -1 | 0 | 1 | 0.7 |
| $\sigma^{\pm}_8$ | 0.25 | 0.75 | 0.04 | 0.7/0.9 | -1 | 0 | 1 | 0.7 |
| $w^{\pm}_0$ | 0.25 | 0.75 | 0.04 | 0.8 | -1.2/-0.8 | 0 | 1 | 0.7 |
| $w^{\pm}_a$ | 0.25 | 0.75 | 0.04 | 0.8 | -1 | -0.2/0.2 | 1 | 0.7 |
| $n^{\pm}_s$ | 0.25 | 0.75 | 0.04 | 0.8 | -1 | 0 | 0.95/1.05 | 0.7 |
| $h^{\pm}$ | 0.25 | 0.75 | 0.04 | 0.8 | -1 | 0 | 1 | 0.65/0.75 |

### 2.5.2 Simulating the CFHTLenS data

In order to understand the statistical properties of the data we have generated a large set of simulated CFHTLenS skies. These mock data were generated from ray-tracing through $N$-body simulations. We used the zHORIZON simulations, performed on the zBOX-2 and zBOX-3 supercomputers at the University of Zürich, described in detail in Smith (2009). Each of the zHORIZON simulations was performed using the publicly available Gadget-2 code (Springel, 2005), and followed the nonlinear evolution under gravity of $N = 750^3$ equal-mass particles in a comoving cube of length $L_{sim} = 1500\,h^{-1}\mathrm{M}pc$; the softening length was $l_{soft} = 60\ h^{-1}\mathrm{k}pc$. For all realizations 11 snapshots were output between redshifts $z = [0, 2]$; further snapshots were at redshifts $z = \{3, 4, 5\}$. The transfer function for the simulations was generated using the publicly available cmbfast code (Seljak & Zaldarriaga, 1996), with high sampling of the spatial frequencies on large scales. Initial conditions were set at redshift $z = 50$ using the serial version of the publicly available 2LPT code (Scoccimarro, 1998; Crocce et al., 2006). The simulations correspond to several cosmological models, with parameters varying around a fiducial model. The latter closely matched the results of the WMAP experiment (Komatsu et al., 2009). There are 8 simulations of the fiducial model, and 4 of each variational model, matching the random realization of the initial Gaussian field with the corresponding one from the fiducial model. Table 2.2 summarizes the cosmological parameters that we simulated.

From each zHORIZON simulation, 16 large fields of view were generated by choosing different observer positions within the simulation box. These large fields have side lengths of $12\,\mathrm{deg}$ and are covered by a regular mesh of $4096^2$ pixels. For each pixel, a light ray was traced back

through the simulation by a multiple-lens-plane algorithm (see Hilbert et al., 2007, 2009), and its distortion due to gravitational lensing was recorded for a set of 45 source redshifts between 0 and 4.

Each of the large fields was used to create four simulated mock-CFHTLenS wide field source galaxy catalogues for each of the different CFHTLenS fields W1 to W4 (i.e., 64 mock catalogues per CFHTLenS field and ZHORIZON simulation). The basis for the simulated source galaxy catalogues are the actual CFHTLenS source catalogues, from which the angular positions and redshifts were taken. The reduced shear $g \equiv \gamma/(1-\kappa)$ for each galaxy in the mock catalogues was computed by multi-linear interpolation of the simulated lensing distortions onto the source galaxy's angular position and redshift (using a different angular offset within the $12 \times 12\,\mathrm{deg}^2$ simulated fields for each mock catalogue). The 'observed' source galaxy ellipticities in the simulated catalogues were then computed by combining the reduced shear from the ray-tracing and the randomly rotated observed ellipticities from the actual source galaxy catalogue.

### 2.5.3 Maps from the CFHTLenS data

As a first step in our analysis of the CFHTLenS data, we compute several aperture based maps for the four fields of the CFHTLenS. To generate these maps the survey area was pixelated and the corresponding maps were computed for apertures located at the pixel centers. We furthermore only include apertures that are at least 20 per cent complete, the map values for all less complete apertures are set to the minimum value and therefore appear as blue pixels. Images of the official CFHTLenS masks are presented in Appendix 2.B.

In Figure 2.3 we show the aperture mass map and its corresponding signal-to-noise map for the W1 field. The aperture masses are estimated using Eq. (2.40) while the noise for each aperture was estimated as (Hetterscheidt et al., 2005):

$$\sigma^2\left(\widehat{M_{\rm ap}(\boldsymbol{\theta}_0|\vartheta)}\right) = \frac{(\pi\vartheta^2)^2\sum_i w_i^2 Q_i^2 \epsilon_{{\rm t},i}^2}{2\left(\sum_i w_i\right)^2}\,. \tag{2.52}$$

The results are shown for the Schneider filter Eq. (2.11) with aperture extent in the set $\vartheta \in [5', 10', 20']$. It is interesting to note that near the survey mask boundaries the value of the aperture mass obtains large positive and negative values. This is due to the fact that for incomplete apertures the ring averaged tangential shear becomes biased as the mask induces non-vanishing $B$-modes (cross-shear). How to deal with the effects of masked apertures will be discussed in detail below and in a companion paper.

In Figure 2.4 we show the aperture completeness map as well as a map of the aperture weights derived from Eq. (2.49), using the shot noise dominated limit of Eq. (2.38). We rescaled the latter

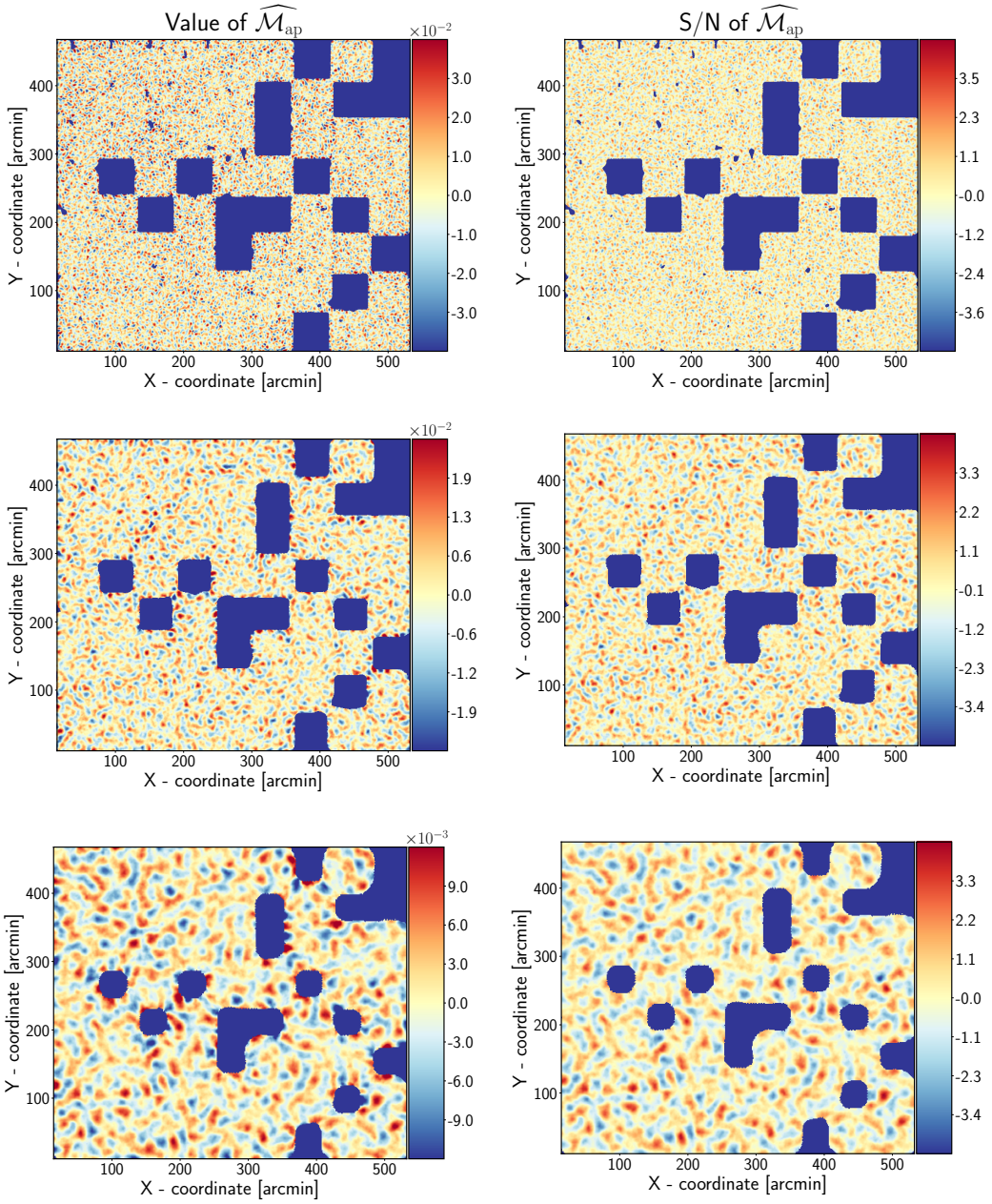


Figure 2.3 Mass related maps measured from the W1 field of the CFHTLenS data. The x- and y-axes show the right-ascension and declination of the survey, in arcminutes. The left and right columns show the results for the aperture mass ($\mathcal{M}_{\rm ap}(\boldsymbol{\theta}_0|\vartheta)$) and the corresponding signal-to-noise ratio. The top, middle and bottom rows show the results for aperture radii of $5'$, $10'$, and $20'$, respectively. The value of the field at each location is indicated by the colour bar on the right.

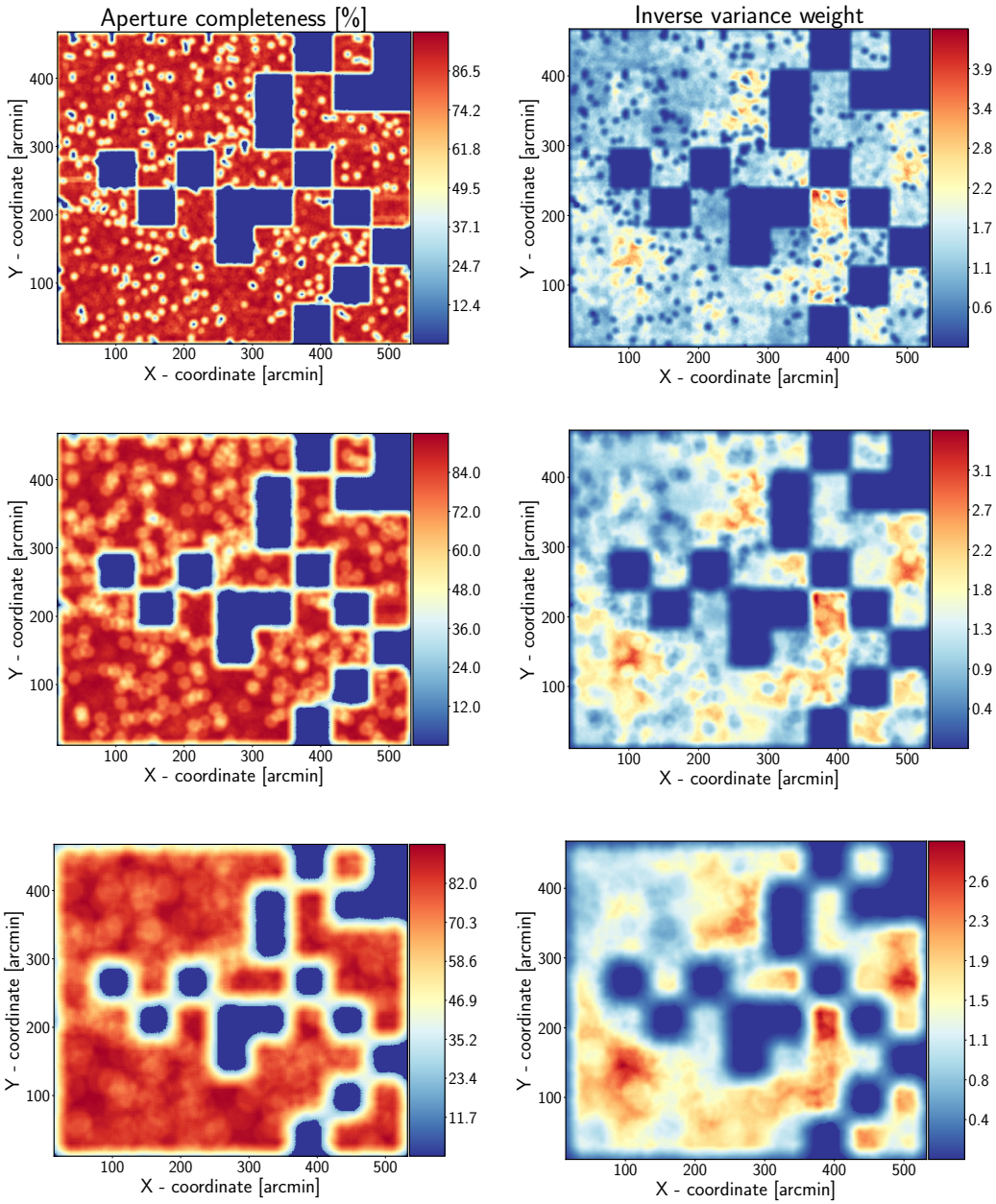


Figure 2.4 Same as Figure 2.3, but this time columns indicate aperture completeness (left) and the aperture weights derived from Eq. (2.49), using the shot noise dominated limit of Eq. (2.38) (right). The weight maps shown are rescaled by their inverse mean. As expected, the weight depends on the local galaxy number density and on the aperture completeness.

map by its inverse mean such that the mean weight becomes unity. From there we can explicitly see that such aperture weights depend on the aperture completeness, cosmic structures and on the local depth of the survey. Analogous maps for the other three fields are presented in Appendix 2.C.

## 2.6 Measurements of the aperture mass variance

We now turn to the estimation of the aperture mass variance from the CFHTLenS data using the direct and correlation function estimators.

### 2.6.1 Analysis of the CFHTLenS mock skies

Before performing our statistical analysis of the real CFHTLenS data we first make a study of the direct and correlation function estimators as applied to the mocks. From this we will be able to determine whether the methods are consistent with one another to within the errors and also which of the three weighting schemes given by Eqs. (2.47) and (2.49) provides the better method for estimating $\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\vartheta)$.

The left panel of Figure 2.5 shows the variance of the aperture mass estimated from the mocks, as a function of the aperture radius. For the results presented in this section we used a spacing of $d = 0.25\ \vartheta$ between the apertures. The thin coloured lines show the results from the direct estimator approach where the estimates from individual apertures are combined with equal weight, but where a completeness thresholds $c_k$ has been adopted – this is equivalent to weight scheme $\mathcal{W}_1$ c.f. Eq. (2.47). For example the magenta lines show a conservative case, where only apertures with completeness $c_k \gtrsim 90\%$ are taken, whereas the red lines are the most relaxed where all apertures with $c_k \gtrsim 10\%$ are allowed. The thick dashed line shows the results obtained from the correlation function method, where the pair counts have been measured using the TreeCorr code of Jarvis et al. (2004)[8]. The grey shaded region shows the standard error for the correlation function method and the error bars are the errors on direct estimator. In all cases these were determined from the ensemble of 512 mocks of our fiducial model. The right panel of Figure 2.5 shows the same as the left panel, however the results from the direct estimator have been binned in completeness. Both panels also show the theoretical prediction evaluated via Eq. (2.20).

[8] We computed the shear correlation functions down to $6''$, used 100 logarithmically spaced bins per decade and set the `binslop` parameter to 0.1.

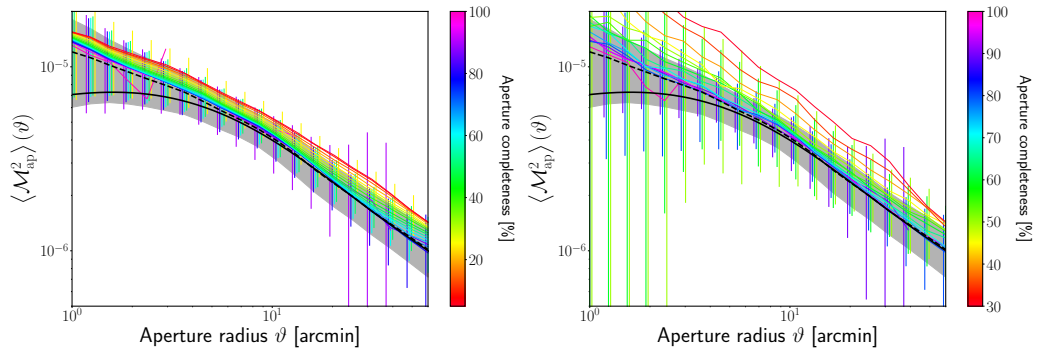


Figure 2.5 *Left hand side:* Aperture mass variance as a function of aperture radius computed from the 512 full-ray traced CFHTLenS mock catalogues of the fiducial runs of the W1 field for the different weighting schemes described in the text. The thick solid line gives the theoretical prediction evaluated via Eq. (2.20). The thin solid lines show results from the direct approach where the estimates from each apertures are combined with equal weight (this is Eq. (2.47) in the text). The colour of the lines indicates the value of the aperture completeness parameter $c_k = A_k/A$ – only apertures with a completeness greater than this value are used in the estimate. The thick dashed line gives the result from the correlation function approach as obtained using the TreeCorr routine (Jarvis et al., 2004). The error bars show the error for a single realisation of the CFHTLenS W1 field. The grey shaded region are for TreeCorr. Note that the errors from the direct estimator approach have been slightly offset for clarity. *Right hand side:* shows the same as the left panel, except this time the estimates from the direct estimator have been computed in bins of aperture completeness. The diffenence between the simulations and the theoretical predictions for small aperture radii can be attributed to shot noise and the line-of-sight discretization from which the simulated data suffers (Hilbert et al., 2020).

The figure shows that there is a clear bias in the direct estimator approach when the apertures have a low completeness. This is manifest as an increased amplitude of the signal on all scales. However, for apertures with $c_k \gtrsim 75\%$ completeness (blue line) we find that the results are in good agreement with the case of $c_k \gtrsim 90\%$ (magenta line) and that these are fully consistent with the correlation function results on large scales, to within the errors. We note that on scales smaller than $\sim 5'$ the results from TreeCorr appear to be biased slightly low to those from the direct estimator approach. We furthermore note a difference between the theoretical prediction and the measurements for small aperture radii which is due to shot noise and the line-of-sight discretization from which the simulated data suffers (Hilbert et al., 2020).

The left panel of Figure 2.6 again shows the aperture mass variance estimated from the mocks for the two methods, but this time the estimates from individual apertures are combined using

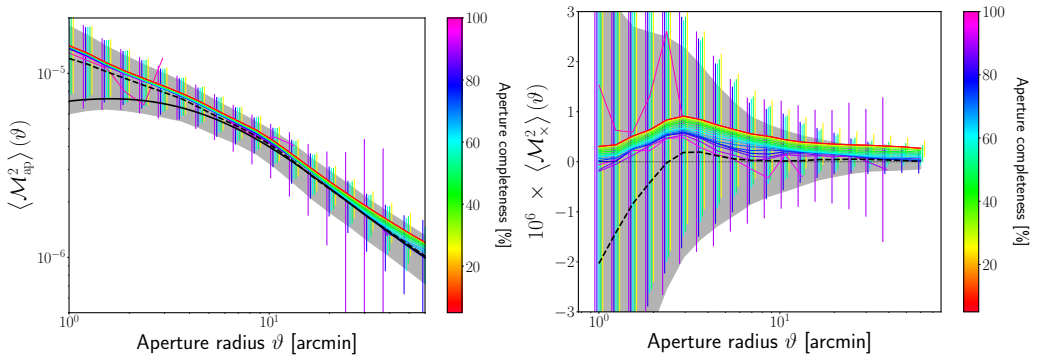


Figure 2.6 *Left hand side:* Same as Figure 2.5, however this time the estimates for each aperture are combined using an inverse variance weighting scheme (this is Eq. (2.49) in the text). Note that when $c_k \lesssim 10\%$ is equivalent to the case where all apertures are considered irrespective of coverage and their results are combined using an inverse variance weight. *Right hand side:* Same as the left panel, but this time showing the variance of the cross component of aperture mass. As shown in Eq. (2.50) and in the absence of systematic errors, $\langle \mathcal{M}_\times^2(\vartheta) \rangle = 0$.

the inverse variance weighting schemes with a completeness threshold – that is we now employ $\mathcal{W}_2$ and $\mathcal{W}_3$, c.f. Eqs. (2.48) and (2.49). The right panel of Figure 2.6 shows the variance of the cross component of the aperture mass $\langle \mathcal{M}_\times^2 \rangle(\vartheta)$ from Eq. (2.50), which for B-mode free fields should vanish.

There are a number of interesting points to note from this analysis. First, we see that all of the estimates from the direct estimator are consistent with one another and that they all lie within the error bars of the TreeCorr result. Nevertheless, for scales $\vartheta > 20'$ we still observe that the direct estimator approach appears to be slightly biased high on large scales for aperture completeness values $c_k \gtrsim 40\%$ compared to the correlation function method. However, for aperture completeness levels $c_k \gtrsim 75\%$ we see excellent agreement between the two methods. On small scales, $\vartheta < 5'$, the correlation function method gives slightly lower results than the direct estimator. Here we believe that the direct estimator is correct, since as was noted in Kilbinger et al. (2006), the correlation function method is biased low on scales of the order $\vartheta \sim 1'$ due to the absence of correlation function bins on small scales. In the TreeCorr code the small-scale cutoff in the pair counts is set at $\vartheta > 6''$. Note that in the mock data there is no image blending and so the direct estimator should not suffer from this suppression. For the observed CFHTLenS data this is not necessarily the case.

We also note that, as shown in the right panel of Figure 2.6, the cross component of the aperture mass variance is consistent with zero to within the error bars for all completeness fractions used. However, the $c_k \gtrsim 40\%$ shows a small positive offset from zero at the level of below

$5 \times 10^{-7}$. On large and small scales the bias is very small for $c_k \gtrsim 75\%$. This gives us further confidence that the discrepancy between the direct estimator and TreeCorr on small scales is due to the bias in the correlation function approach.

Another point of interest, is that we see the error bars on the most conservative completeness cuts, $c_k \gtrsim 95\%$, are significantly larger than those obtained from the correlation function method. This makes sense, since, for the case of the most conservative cuts, one can find only a few apertures that meet the criterion. On the other hand, for completeness fractions of the order $c_k \gtrsim 70\%$, the error bars between the two methods are comparable.

Based on the discussion above, we will be using the weighting scheme $\mathcal{W}_3$ for the analysis of the observed CFHTLenS data.

Finally, we note that the runtime in evaluating the estimators is similar for the direct estimator and TreeCorr, with the former being roughly a factor of two faster. It is also worth noting that TreeCorr is not exact, but is using various (well tested) approximations and has also been highly optimized to reach this speed. As noted earlier, the direct estimator speed also depends on the rate of aperture oversampling. Repeating the analysis for sparser aperture sampling with $d = 0.5\,\vartheta$ we found that (as expected) the runtime was furthermore reduced by a factor of four (to the order of a minute for a W1 field on a single processor) with only a tiny fraction of the signal-to-noise being lost. We postpone a more thorough analysis of the runtime to our companion paper.

### 2.6.2 Calibration of the estimators for ellipticity bias

We now turn to the analysis of the observed data. As discussed in §2.4 the aperture mass variance for a set of apertures can be directly estimated using Eqs. (2.32) and (2.43). However, owing to calibration errors in the *lens*fit shape estimation algorithm (Miller et al., 2013), each galaxy's ellipticity has a corresponding multiplicative $m$ and additive $c$ bias. Hence, the observed and true ellipticity components of the $i$th galaxy are related through:

$$\epsilon_{1,i}^{\rm obs} = (1 + m_i)\epsilon_{1,i}^{\rm true} + c_{1,i}\ ; \tag{2.53}$$

$$\epsilon_{2,i}^{\rm obs} = (1 + m_i)\epsilon_{2,i}^{\rm true} + c_{2,i}\ . \tag{2.54}$$

For the CFHTLenS $c_1$ was found to be consistent with zero, however $c_2$ was found to have a $S/N$ and size dependent bias that was subtracted from each galaxy. On average this correction was of the order $2 \times 10^{-3}$.

**Correction for shear correlation functions:** following Miller et al. (2013), the shear correla-

tion functions can be corrected for multiplicative bias through the following procedure. The 'raw' shear correlation functions are first estimated from the 'observed' shears using the estimator:

$$\widehat{\xi_{\pm}^{\rm raw}}(\theta_\alpha) = \frac{\sum_i \sum_{j\neq i} w_i w_j \left[\epsilon_{{\rm t},i}^{\rm obs}\epsilon_{{\rm t},j}^{\rm obs} \pm \epsilon_{\times,i}^{\rm obs}\epsilon_{\times,j}^{\rm obs}\right] \Pi(\theta_{ij}|\theta_\alpha)}{\sum_i \sum_{j\neq i} w_i w_j \Pi(\theta_{ij}|\theta_\alpha)} \,, \tag{2.55}$$

where $\theta_{ij} \equiv |\boldsymbol{\theta}_i - \boldsymbol{\theta}_j|$ and where $\Pi(\theta_{ij}|\theta_\alpha)$ is the pair binning function which is unity if $\theta_{ij}$ lies in the range $[\theta_\alpha - \Delta\theta/2, \theta_\alpha + \Delta\theta/2)$. If we now insert Eqs. (2.53) and (2.54) into the above estimator (taking $c_{1,i} = 0$ and $c_{2,i} = 0$) we find:

$$\widehat{\xi_{\pm}^{\rm raw}(\theta_\alpha)} = \frac{\sum_{j\neq i} w'_i w'_j \left[\epsilon_{{\rm t},i}^{\rm true}\epsilon_{{\rm t},j}^{\rm true} \pm \epsilon_{\times,i}^{\rm true}\epsilon_{\times,j}^{\rm true}\right] \Pi(\theta_{ij}|\theta_\alpha)}{\sum_{j\neq i} w_i w_j \Pi(\theta_{ij}|\theta_\alpha)} \,, \tag{2.56}$$

where in the above we introduced the new weights $w'_i \equiv w_i(1+m_i)$, then we see that the above equation can be rewritten:

$$\begin{aligned}\widehat{\xi_{\pm}^{\rm raw}(\theta_\alpha)} = &\left(\frac{\sum_{j\neq i} w'_i w'_j \left[\epsilon_{{\rm t},i}^{\rm true}\epsilon_{{\rm t},j}^{\rm true} \pm \epsilon_{\times,i}^{\rm true}\epsilon_{\times,j}^{\rm true}\right] \Pi(\theta_{ij}|\theta_\alpha)}{\sum_{j\neq i} w'_i w'_j \Pi(\theta_{ij}|\theta_\alpha)}\right) \\ &\times \left(\frac{\sum_{j\neq i} w'_i w'_j \Pi(\theta_{ij}|\theta_\alpha)}{\sum_{j\neq i} w_i w_j \Pi(\theta_{ij}|\theta_\alpha)}\right) \,,\end{aligned} \tag{2.57}$$

The first term on the right in parenthesis is the 'calibrated' true shear correlation which we can write as $\xi_{\pm}^{\rm cal}$, hence we can write:

$$\widehat{\xi_{\pm}^{\rm cal}(\theta_\alpha)} = \frac{\xi_{\pm}^{\rm raw}(\theta_\alpha)}{1+\mathcal{K}(\theta_\alpha)} \;\; ; \;\; 1+\mathcal{K}(\theta_\alpha) \equiv \frac{\sum_{j\neq i} w'_i w'_j \Pi(\theta_{ij}|\theta_\alpha)}{\sum_{j\neq i} w_i w_j \Pi(\theta_{ij}|\theta_\alpha)} \,. \tag{2.58}$$

**Correction for direct aperture mass variance estimator:** following in the footsteps of the shear correlation function approach, we define the raw uncorrected aperture mass variance estimate as:

$$\left[\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{\rm raw} = (\pi\vartheta^2)^2 \, \frac{\sum_{j\neq i} w_i \, w_j \, Q_i \, Q_j \, \epsilon_{{\rm t},i}^{\rm obs} \, \epsilon_{{\rm t},j}^{\rm obs}}{\sum_{j\neq i} w_i w_j} \,, \tag{2.59}$$

recalling that the sums only extend over the galaxies in the aperture. If we now insert Eqs. (2.53) and (2.54) into the above estimator, as before, then we find:

$$\left[\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{\rm raw} = (\pi\vartheta^2)^2 \, \frac{\sum_{j\neq i} w'_i \, w'_j \, Q_i \, Q_j \, \epsilon_{{\rm t},i}^{\rm true} \, \epsilon_{t,j}^{\rm true}}{\sum_{j\neq i} w_i w_j} \,. \tag{2.60}$$

Again, on redefining the *lens*fit weights this leads us to write:

$$\begin{aligned}\left[\widehat{M_{\rm ap}^2(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{\rm raw} = &\left((\pi\vartheta^2)^2 \frac{\sum_{j\neq i} w'_i \, w'_j \, Q_i \, Q_j \, \epsilon_{{\rm t},i}^{\rm true} \, \epsilon_{{\rm t},j}^{\rm true}}{\sum_{j\neq i} w'_i w'_j}\right) \\ &\times \left(\frac{\sum_{j\neq i} w'_i w'_j}{\sum_{j\neq i} w_i w_j}\right) \,.\end{aligned} \tag{2.61}$$

Thus we see that the calibrated estimate of the aperture mass variance can be written:

$$\left[\widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{\rm cal} = \frac{\left[\widehat{M^2_{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}\right]^{\rm raw}}{1+\mathcal{K}^{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k})}\,, \tag{2.62}$$

where the normalisation factor is

$$\begin{aligned} 1+\mathcal{K}^{\rm ap}(\vartheta|\boldsymbol{\theta}_{0,k}) &\equiv \frac{\sum_{j\neq i} w_i w_j (1+m_i)(1+m_j)}{\sum_{j\neq i} w_i w_j} \\ &= \frac{\langle w(1+m)\rangle^2}{\langle w\rangle^2}\frac{1-S_2^{(1+m)}}{1-S_2}\,, \end{aligned} \tag{2.63}$$

where, in the same vein as in §2.4.2, we defined

$$S_2^{(1+m)} \equiv \frac{\langle w^2(1+m)^2\rangle}{\langle w(1+m)\rangle^2}\,. \tag{2.64}$$

### 2.6.3 Analysis of the CFHTLenS data

In left panel of Figure 2.7 we show the aperture mass variance and in the right the variance of the cross component variance, estimated from the CFHTLenS. The results shown are for the full combination of the W1, W2, W3 and W4 survey areas. For the direct aperture mass variance approach the estimate was obtained using the weighting scheme $\mathcal{W}_3$, c.f. Eq. (2.49), and the error was obtained using a weighted jackknife resampling of the data in patches having an area of roughly $(1\ \deg)^2$. For the method of measuring the aperture mass variance from the shear correlation functions (thick, black dashed line), when combining the results from each field, we have simply weighted the estimates in proportion to the field area.

In Fig. 2.7 the vertical dashed line indicates the scale $\vartheta = 5.5'$ identified in Kilbinger et al. (2013), above which the E/B-mode leakage is claimed to be $\lesssim 1.5\%$. The E/B mode leakage on smaller scales originates from the $9''$ cut-off in the shear-correlation functions, which originates from the blending of galaxy images and a shear bias for close pairs (see §2.6.4).

The important point to note from the figure is that on scales $\vartheta > 5.5'$ the measurements from the direct and correlation function estimators are in very good agreement to within the errors for a wide range of aperture completeness thresholds. In addition, with the exception of all but the lowest $c_k$ thresholds, the variance of the cross component of the aperture mass variance is consistent with zero on the same scales. On scales $\vartheta \leq 5.5'$, both estimators appear to have non-zero B-modes, which rise sharply on scales $\vartheta \lesssim 2'$.

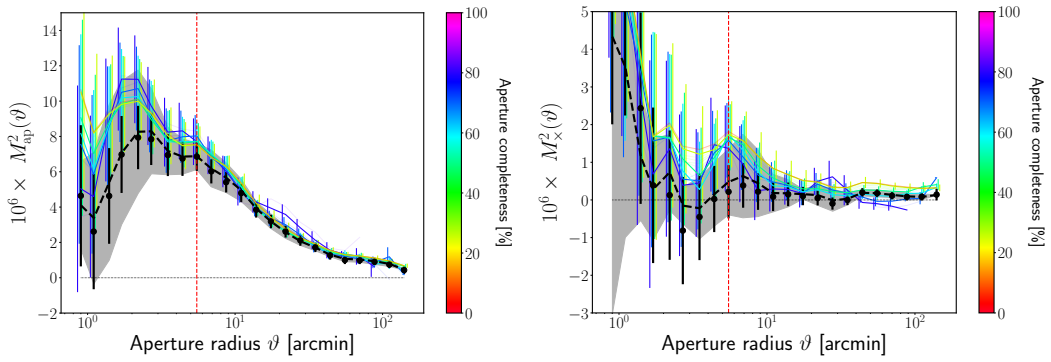


Figure 2.7 *Left hand side:* Aperture mass variance as a function of aperture cut-off scale as measured in the CFHTLenS data. The results for the direct estimator are indicated as coloured lines, where the line colour indicates the value of aperture completeness that was used. The thick dashed line gives the result from the correlation function approach as obtained using the TreeCorr routine (Jarvis et al., 2004). The error bars show the error for the full CFHTLenS. *Right hand side:* : Same as the left panel, but for the variance of the cross component of the aperture mass.

### 2.6.4 Impact of close pair image blending on the variance

As noted in Miller et al. (2013), *lens*fit galaxies with separations closer than $9''$ tend to have a bias in their ellipticities which is in the direction of the line connecting the centres of the galaxy images. For the shear correlation functions, Kilbinger et al. (2013) stated that this bias can be removed by only computing the shear correlation functions down to $9''$. Their justification was that the alignment direction of the pair is, to a very good approximation, randomly oriented. We now investigate how such a bias may contaminate the direct estimator approach for aperture mass variance by comparing the measurements of the full catalogue to a reduced one in which clustered galaxies below the blending scale are removed.

We proceed by describing our algorithm for removing close pairs of galaxies to generate a reduced catalogue.

1. We begin by initializing a Boolean mask value for each galaxy. These values are all initially set to unity.

2. We then spatially organize the data using a hierarchical KD-tree.

3. We next set the pair-cut-off scale $\theta_{\rm pc}$, within which galaxy pairs are to be expunged from the catalogue.

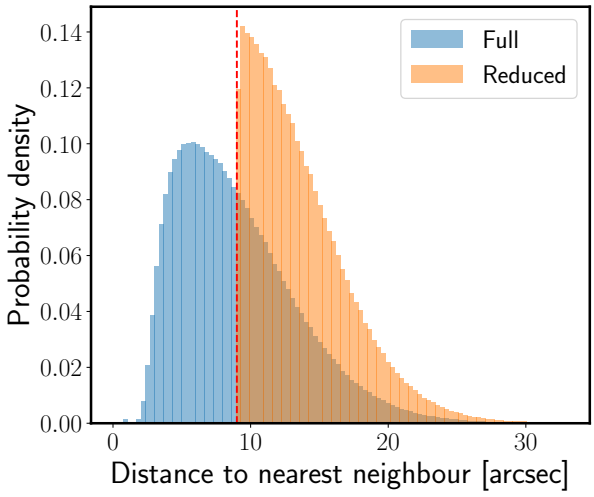


Figure 2.8 The probability density functions of the nearest neighbour distances in the CFHTLenS. The blue histogram shows the distribution before applying our exclusion algorithm and the orange one shows the result after setting the pair-cut-off scale $\theta_{\rm pc} = 9''$. The vertical dashed red line indicates $\theta_{\rm pc}$.

4. We now loop over all galaxies and perform a range search. If any of the galaxy positions lie within the sphere of radius $\theta_{\rm pc}$ and have their Boolean flag set to unity, then the Boolean mask value associated with the current galaxy is set to zero and the galaxy will henceforth be excluded when estimating statistics from the data.

We apply this method to the CFHTLenS data and set the pair-cut-off scale $\theta_{\rm pc} = 9''$. This yields a reduced catalogue that contains $\sim$70% of the original galaxies.

Figure 2.8 shows the probability density function of the distance to the nearest-neighbour galaxy. Note that these curves have been normalised so that the area under the graph gives unity.

Figure 2.9 shows our measurements of the aperture mass variance and the variance of the cross component of aperture mass, in the full and the close-pair reduced CFHTLenS shear catalogues. We compare the results from both the direct and the correlation function estimators. For the direct estimator approach we have employed the weighting function $\mathcal{W}_3$, with the completeness threshold set to $c_k = 0.7$. Some important points can be noted: looking at the left panel of Fig. 2.9 the result of excluding pairs of galaxies that are closer then 9" lowers the amplitude of $\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle$ by less than 15% for small aperture scales ($\sim 10'$). For larger apertures the changes are smaller. Interestingly, the results from the TreeCorr analysis of the full and reduced catalogue also show similar differences: for small aperture scales, the aperture mass variance

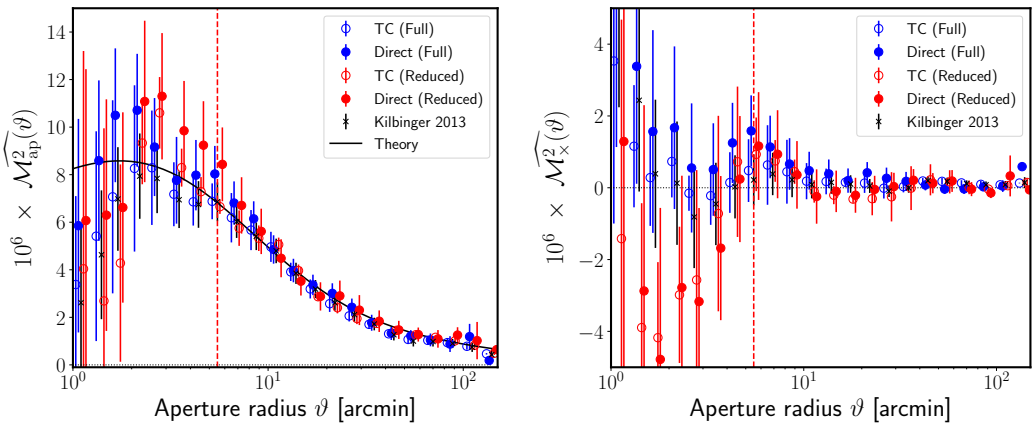


Figure 2.9 *Left hand side*: Comparison of the aperture mass variance from the full and small scale-pair reduced CFHTLenS catalogues. The open circles show the results from TreeCorr and the solid points the results from the direct estimator when including all apertures with $c_k > 0.7$. The red and blue colours indicate the full and reduced catalogues, respectively. The black crosses and the black line show the published measurements from Kilbinger et al. (2013) and the corresponding theoretical prediction for their best-fit cosmology evaluated via Eq. (2.20). The vertical red dashed line indicates the scale above which the E/B mode leakage should be less than 1% for the correlation function method. *Right hand side*: same as left panel, except this time for the cross component of the aperture mass variance.

appears to be lower in the reduced catalogue. In addition, we see that all of the estimators agree to within the errors on all scales. However, the agreement between the direct estimator and the TreeCorr result is exceptionally good for the measurements from the reduced catalogue.

It is also interesting to compare these results with the published measurement from Kilbinger et al. (2013) (denoted as the black crosses in the plot). Here we see that our measurements are fully consistent with the Kilbinger et al. (2013) measurements, to within the errors. The right panel of Figure 2.9 shows that the variance of the cross-component of the aperture mass is reassuringly consistent with zero for both estimators applied to both the full and reduced catalogues.

## 2.7 Information content of the estimators

We now turn our attention to the question of addressing the possible information loss in using the direct estimator approach. We will do this using the Fisher matrix formalism. If we assume that the likelihood function for measuring the aperture mass variance for a set of $N_d$ aperture scales is Gaussian, and that the priors on the cosmological parameters are flat, then the Fisher

information matrix can be written (Tegmark et al., 1997):

$$\mathcal{F}_{\alpha\beta} = \frac{1}{2}\mathrm{T}r\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial p_\alpha}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial p_\beta}\right] + \frac{\partial \boldsymbol{\mu}^{\mathrm{T}}}{\partial p_\alpha}\mathbf{C}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial p_\beta}\,, \tag{2.65}$$

where $\boldsymbol{\mu}^{\mathrm{T}} = \left(\widehat{\mathcal{M}^2_{\mathrm{ap}}(\vartheta_1)}, \ldots, \widehat{\mathcal{M}^2_{\mathrm{ap}}(\vartheta_{N_d})}\right)$ is the set of model means measured at the bin centres and $\mathbf{C}$ is the model covariance matrix. The vector $\boldsymbol{p}$ is the set of cosmological parameters. In this study we will restrict our attention to the cosmological parameters $\sigma_8$ and $\Omega_{\mathrm{m},0}$, as these are the most readily constrained from lensing data. The minimum variance bounds on a given cosmological parameter, after marginalising over all other parameters, can be obtained as:

$$\sigma^2_{p_\alpha} = \left[\mathcal{F}^{-1}\right]_{\alpha\alpha}\,. \tag{2.66}$$

In order to simplify the calculation, we will assume that the first term on the right-hand-side of Eq. (2.65) is significantly smaller than the other term. As was noted in §6 of Smith & Marian (2015) this can be justified in the high-$k$ limit through mode counting arguments, however, it is in general an incorrect assumption[9].

In order to compute the second term of Eq. (2.65) we follow the approach lain down in Smith et al. (2014) and use the mocks to evaluate all quantities. That is we measure the derivatives of the model mean with respect to the cosmological parameters and we also estimate the precision matrix $\mathbf{C}^{-1}$. To do these tasks we make use of the ray traced mock CFHTLenS data. The derivatives are estimated using:

$$\widehat{\frac{\partial \mu_i}{\partial p_\alpha}} = \sum_{j=1}^{N_{ens}} \frac{\widehat{\mathcal{M}^2_{\mathrm{ap},(j)}}(\vartheta_i|p_\alpha + \Delta p_\alpha) - \widehat{\mathcal{M}^2_{\mathrm{ap},(j)}}(\vartheta_i|p_\alpha - \Delta p_\alpha)}{2N_{ens}\Delta p_\alpha}\,, \tag{2.67}$$

where $N_{ens}$ is the number of realisation of the ensemble, and $\mathcal{M}^2_{\mathrm{ap},(j)}(\theta_i|p_\alpha + \Delta p_\alpha)$ is the estimate of the aperture mass variance on scale $\theta_i$ in the $j$th realisation of the mocks, for the simulation with cosmological parameters $p_\alpha + \Delta p_\alpha$. In computing Eq. (2.67) we use the 256 mock ray-tracing simulations of each cosmological variation of the CFHTLenS data. Note that the above estimator will reduce the cosmic variance, since when running the modified cosmology simulations we used Fourier phase realisations that were identical to the fiducial model runs. In addition, when estimating the precision matrix we take account of the bias in the estimator using the method described in Hartlap et al. (2007):

$$\widehat{\mathbf{C}^{-1}} = \frac{N_{ens} - N_{\mathrm{d}} - 2}{N_{ens} - 1}\left[\widehat{\mathbf{C}}\right]^{-1}\,, \tag{2.68}$$

[9] On the other hand, as was discussed in Takahashi et al. (2011) the likelihood is not Gaussian, since the power spectrum estimator is $\chi^2$ distributed. Thus one should not quote over-optimistic errors based on the wrong form for the likelihood function.

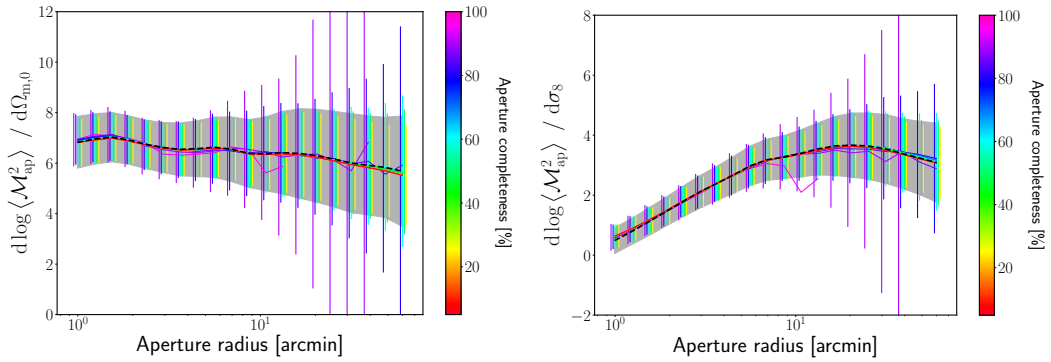


Figure 2.10 Estimates of the derivatives of the aperture mass variance with respect to the cosmological parameters. The left and right panels show the results for the variations with respect to the matter density parameter $\Omega_{\rm m,0}$ and the variance of matter fluctuations $\sigma_8$, respectively. The solid lines all show the results measured from the direct estimator approach, where the colour of the line indicates the value for the aperture completeness parameter that was adopted. The error bars indicate the errors in the ensemble. The black dashed line shows the results from the correlation function approach as measured using TreeCorr. The shaded region shows the standard error on the TreeCorr estimates.

where $N_{\rm d}$ is the dimension of the data vector and $\widehat{\mathbf{C}}$ is the standard maximum likelihood estimator of the covariance matrix of the data. We note that by applying the Hartlap correction to our estimate of the covariance matrix does not account for the sampling distribution of of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\mathbf{C}^{-1}}$ which can be found in Jeffrey & Abdalla (2019); their resulting likelihood would lead to a more sharply peaked center and broader tails and it would need to be propagated in the information theoretic definition of the Fisher matrix to derive the analog of (2.65) in a similar vein to Sellentin & Heavens (2017).

Figure 2.10 shows the derivatives of the aperture mass variance with respect to the cosmological parameters, estimated from the 256 ray-traced mock simulations. The agreement between the results from the direct estimator approach is excellent, for all of the $c_k$ values that we considered, where we used the inverse variance weighting approach of $\mathcal{W}_3$ to combine estimates from individual apertures. In addition, these also agree with the results from the correlation function approach to a high degree of accuracy. The only issue to note is that for high completeness fractions $c_k \sim 1$ the estimates become very noisy[10].

Figure 2.11 shows the 2D confidence contour for $\Omega_{\rm m,0}$ and $\sigma_8$, and also the 1D posterior distributions, marginalised over all other parameters, for the same two parameters. Focusing on the

[10] In making the Fisher forecasts we need to take into account the error on the model means and the precision matrix, since not doing so could lead to over-optimistic forecasts. This could be done by generalising the approach described in Taylor et al. (2013).

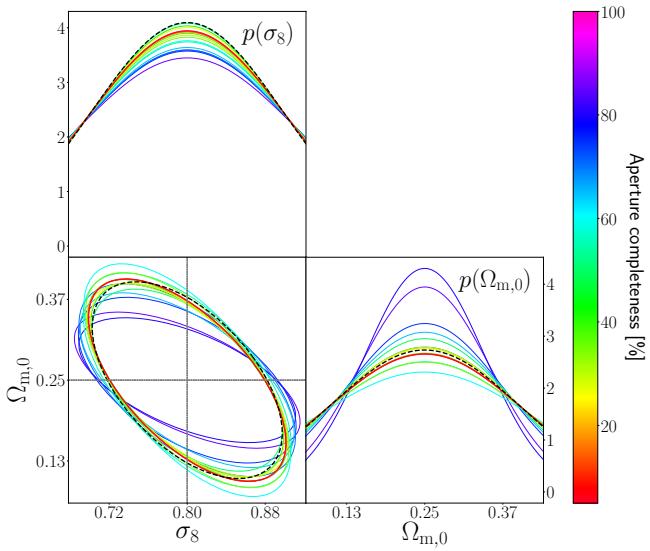


Figure 2.11 Top left and bottom right panels show the 1D posterior distributions, marginalised over all other parameters, for $\Omega_{m,0}$ and $\sigma_8$, respectively. The bottom left panel shows the 2D confidence interval ellipse for the joint posterior distribution of $\Omega_{m,0}$ and $\sigma_8$. The coloured lines again show the results from the Fisher matrix forecast using the direct estimator approach for various aperture completeness thresholds $c_k$. The dashed line shows the results from TreeCorr.

2D contours, we notice a number of interesting points: first, considering the results for the direct estimator, we see that as we decrease the threshold completeness value $c_k$, the ellipses rotate clockwise by a small amount. This difference in orientation can be explained by noting that for a CFHTLenS like survey the off-diagonal elements of the covariance matrix are getting more and more noisy for increasing values of $c_k$. The reason for this is that the effective survey footprint changes for different aperture radii and that this difference is most prominent for conservative aperture completeness cuts. When choosing a $c_k$ of around $0.7$ or lower, the covariance matrices become stable and so does the orientation and area of the ellipse.

Similar observations can be made for the 1D marginalized posterior distributions which lets us conclude that, within the Fisher Matrix formalism, the information content of the direct estimator is comparable with the correlation function method.

There are some caveats that must be mentioned to the interpretation of these results. Firstly, both the precision matrix and the model means are estimated from the mock simulations of the CFHTLenS, and are therefore subject to errors. This means that the forecasted constraining power should also come with an error. Since we are only interested in the relative information, we have not taken this into account. A more detailed study is required in order to make a more precise statement, and this is beyond the scope of this work.

## 2.8 Conclusions and discussion

In this paper we have explored an alternative method for estimating the aperture mass statistics in weak lensing cosmic shear surveys. Our method is a direct estimator of the variance. With the use of a hierarchical KD-tree algorithm for ordering the data we found that the computational time for execution of this estimator was linear in the number of galaxies per aperture and the number of apertures used in the estimate. This paper, the first in a series, focused on the two-point statistics, and in particular the aperture mass variance. The summary is:

In §2.2 and §2.3 we reviewed the background theory of weak lensing in the cosmological context and the aperture mass variance and its connection to the matter power spectrum. Here we also discussed the standard approach for estimating this quantity, which relies on measurements of the shear correlation functions.

In §2.3 we introduced the direct estimator for the aperture mass variance that we employed. We showed that when including ellipticity weights the estimator is unbiased. We also computed the variance of the estimator and showed that in the limit a of Gaussian shear signal, no source clustering and a large number of galaxies per aperture, the variance reduces to a simple expression. We then showed how the original estimator introduced by Schneider et al. (1998) can be accelerated to linear order in the number of galaxies per aperture. We also discussed various weighting schemes for combining the estimates from different apertures. Finally, we illuminated the computational complexity of our estimator.

In §2.4 we gave an overview of the CFHTLenS data and we also described our method for generating mocks of the survey using full gravitational ray-tracing simulations through $N$-body simulations. As a first test, we measured the aperture mass maps from the survey data.

In §2.6 we computed the aperture mass variance from our mock surveys, using both the direct estimator and correlation function method. We found that if we included incomplete apertures in the direct estimator method, and if we combined all apertures equally that there was a significant bias in the final result. This bias vanished if only complete apertures were used. However, the errors on the estimates increased significantly. We then explored an alternative weighting scheme, where the apertures were combined using an inverse variance weighting approach, where the variance was assumed to be dominated by shape noise. The results in this case were found to be in excellent agreement with the alternative method, even in the case where incomplete apertures were included in the estimate. We found that an aperture completeness threshold of $\sim 0.7$ gave very good results and contained only a small residual of B-modes, that were contained well

within the error tolerance.

We then turned to the application of the method to the CFHTLenS data. It was necessary to account for two additional observational biases: first, we derived the correction factors required to account for the ellipticity bias in our estimator; second, we created a modified source galaxy catalogue that removed pairs of galaxies whose images were in close projection on the sky whose ellipticities are biased by an artefact in the ellipticity estimator algorithm *lens*fit. On taking account of these we found that our direct estimator approach and our estimates using the shear correlation function were in excellent agreement with the published data from Kilbinger et al. (2013).

In §2.7 we explored the information content of the direct estimator approach and compared it with that from the correlation function method. We found that the 1D marginalised posterior distributions for $\sigma_8$ were less constraining for high aperture completeness than the shear correlation function method, but that as some incomplete apertures were included in the estimate the distributions became very similar. This trend was mirrored for $\Omega_{\mathrm{m},0}$, except that for high completeness the distribution for the direct estimator was the most constraining. This may be due to errors in the forecast due to uncertainties in the precision matrix and derivatives. The 2D confidence contours for $\Omega_{\mathrm{m},0}$ and $\sigma_8$ were roughly the same size for all aperture completeness thresholds, however they rotated to be in the same direction as those of the correlation function approach as lower thresholds were taken. This leads us to conclude, at this stage, that the information content of the two estimators is comparable.

The main advantage of this development in not to replace the correlation function approach as *the way* to measure aperture mass statistics and other associated statistics, but to show that it is a credible method. As mentioned earlier, the real advantage of this approach is that it can be easily generalised to enable the measurement of higher order aperture mass statistics such as the skewness and kurtosis with very little extra effort. While these can of course also be estimated using the shear three-point and four-point correlation functions, the task of measuring these correlation functions and all of their configurations becomes increasingly onerous and time consuming. The method that we have developed will scale linearly with the number of galaxies in the aperture. This will be the subject of our upcoming study. Besides this, for application to more recent large-scale lensing surveys like DES and KiDS the analysis will need to be extended to include curved sky effects and we see no obvious issues with this.

## Acknowledgements

We thank Gary Bernstein and the anonymous referee for useful comments on the draft.
LP acknowledges support from a STFC Research Training Grant (grant number ST/R505146/1). RES acknowledges support from the STFC (grant number ST/P000525/1, ST/T000473/1). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. This research made use of `numpy`, a library used for scientific computing and technical computing and `matplotlib`, a Python library for publication quality graphics (Oliphant, 2006; Hunter, 2007).
This work was finalized during the Covid-19 outbreak. The authors would like to thank all essential workers around the world that continue to make huge sacrifices to overcome this pandemic.

## Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

# Appendix

## 2.A Supporting calculations for chapter 5.4

### Unbiasedness of the direct estimator

We show that (2.32) provides an unbiased estimator for the true aperture mass dispersion. This can be done through applying three averaging processes: averaging over the intrinsic ellipticity distributions $A$; then the source galaxy positions $P$; and then the ensemble average over the cosmic fields $E$ (following the notation of Schneider et al., 1998). Ignoring the prefactor and the denominator for a moment, if we perform the $A$ average then we get:

$$\begin{aligned}
A\left(\sum_{i\neq j} w_i w_j Q_i Q_j \epsilon_{\mathrm{t},i}\epsilon_{\mathrm{t},j}\right) &= A\left(\sum_{i,j} w_i w_j Q_i Q_j \epsilon_{\mathrm{t},i}\epsilon_{\mathrm{t},j}\right) - A\left(\sum_i w_i^2 Q_i^2 \epsilon_{\mathrm{t},i}^2\right) \\
&= \sum_{i,j} w_i w_j Q_i Q_j \left\{\gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j} + 2\gamma_{\mathrm{t},i}\, A\left(\epsilon_{\mathrm{t},j}^{(s)}\right) + A\left(\epsilon_{\mathrm{t},i}^{(s)}\epsilon_{\mathrm{t},j}^{(s)}\right)\right\} \\
&\quad - \sum_i w_i^2 Q_i^2 \left\{\gamma_{\mathrm{t},i}^2 + 2\,\gamma_{\mathrm{t},i}\, A\left(\epsilon_{\mathrm{t},i}^{(s)}\right) + A\left([\epsilon_{\mathrm{t},i}^{(s)}]^2\right)\right\} \\
&= \sum_{i,j} w_i w_j Q_i Q_j \left\{\gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j} + \frac{\sigma_\epsilon^2}{2}\delta_{ij}^K\right\} - \sum_i w_i^2 Q_i^2 \left\{\gamma_{\mathrm{t},i}^2 + \frac{\sigma_\epsilon^2}{2}\right\} \\
&= \sum_{i\neq j} w_i w_j Q_i Q_j \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\ . \qquad (2.69)
\end{aligned}$$

Note that in the above we assumed that each galaxies' intrinsic ellipticity is indiviually drawn from the same Gaussian distribution $\mathcal{G}\left(0, \sigma_\epsilon^2\right)$ with zero mean and the shape noise $\sigma_\epsilon^2$ as variance, i.e. no intrinsic alignments. Next, we perform the average over the spatial positions of the source galaxies:

$$P\left(\sum_{i\neq j} w_i w_j Q_i Q_j \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\right) = \prod_{\alpha=1}^{N}\left\{\frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2}\right\}\sum_{i\neq j} w_i w_j Q_i Q_j \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}$$

$$= \frac{\sum_{i\neq j} w_i w_j}{(\pi\vartheta^2)^2} \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \mathrm{d}^2\boldsymbol{\theta}_2 \; Q_1 Q_2 \gamma_{\mathrm{t},1}\gamma_{\mathrm{t},2} \; . \tag{2.70}$$

In the first step we took the joint PDF of spatial positions to be simply the product of the independent 1-point PDFs for a uniform random distribution. In the second step, on noting that $\gamma_{\mathrm{t},i} = \gamma_{\mathrm{t},i}(\boldsymbol{\theta}_i;\boldsymbol{\theta}_0)$, where $\theta_0$ is the same for all the galaxies, we used the fact that the spatial integral will yield the same result no matter of the indices - hence the change $(i,j) \to (1,2)$. In the last step, we integrated out the remaining PDFs and rewrote the domain. Finally, we perform the expectation over the cosmic fields:

$$\begin{aligned} E\left\{P\left[A\left(\widehat{M^2_{\mathrm{ap}}(\vartheta)}\right)\right]\right\} &= \frac{(\pi\vartheta^2)^2}{\sum_{i\neq j} w_i w_j} \frac{\sum_{i\neq j} w_i w_j}{(\pi\vartheta^2)^2} \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \mathrm{d}^2\boldsymbol{\theta}_2 \; Q_1 Q_2 \left\langle \gamma_{\mathrm{t},1}\gamma_{\mathrm{t},2}\right\rangle \\ &= \left\langle \mathcal{M}^2_{\mathrm{ap}} \right\rangle \; . \end{aligned} \tag{2.71}$$

### Estimating the variance of the estimator – including the source weights

We can also calculate the variance of the estimator:

$$\begin{aligned} \mathrm{Var}\left[\widehat{M^2_{\mathrm{ap}}(\vartheta)}\right] &= \left\langle (\pi\vartheta^2)^2 \frac{\sum_{i\neq j} w_i \; w_j \; Q_i \; Q_j \; \epsilon_{\mathrm{t},i} \; \epsilon_{\mathrm{t},j}}{\sum_{i\neq j} w_i w_j} (\pi\vartheta^2)^2 \frac{\sum_{k\neq l} w_k \; w_l \; Q_k \; Q_l \; \epsilon_{\mathrm{t},k} \; \epsilon_{\mathrm{t},l}}{\sum_{k\neq l} w_k w_l} \right\rangle \\ &\quad - \left\langle \mathcal{M}^2_{\mathrm{ap}} \right\rangle^2 \\ &= \frac{(\pi\vartheta^2)^4}{\left(\sum_{i\neq j} w_i w_j\right)^2} \sum_{i\neq j} \sum_{k\neq l} w_i w_j w_k w_l Q_i Q_j Q_k Q_l \left\langle \epsilon_{\mathrm{t},i}\epsilon_{\mathrm{t},j}\epsilon_{\mathrm{t},k}\epsilon_{\mathrm{t},l}\right\rangle - \left\langle \mathcal{M}^2_{\mathrm{ap}} \right\rangle^2 \end{aligned} \tag{2.72}$$

Following the recipe as for the mean estimator, we first calculate the average over the source galaxies and this yields Schneider et al. (1998):

$$\begin{aligned} A\left[\epsilon_{\mathrm{t},i}\epsilon_{\mathrm{t},j}\epsilon_{\mathrm{t},k}\epsilon_{\mathrm{t},l}\right] = \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k}\gamma_{\mathrm{t},l} + A\left[\epsilon^{(\mathrm{s})}_{\mathrm{t},i}\epsilon^{(\mathrm{s})}_{\mathrm{t},j}\epsilon^{(\mathrm{s})}_{\mathrm{t},k}\epsilon^{(\mathrm{s})}_{\mathrm{t},l}\right] + \frac{\sigma^2_\epsilon}{2}\Big[\gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\delta^K_{kl} + \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},k}\delta^K_{jl} \\ + \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},l}\delta^K_{jk} + \gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k}\delta^K_{il} + \gamma_{\mathrm{t},j}\gamma_{\mathrm{t},l}\delta^K_{ik} + \gamma_{\mathrm{t},k}\gamma_{\mathrm{t},l}\delta^K_{ij}\Big] \; . \end{aligned} \tag{2.73}$$

Let us now work out the first term on the right-hand-side of the above expression. We see that we have the following possibilities: all indices different $(i \neq j \neq k \neq l) \to$ (1a); two indices equal and three not $(i \neq j \neq k = l) + \mathrm{perms} \to$ (1b); $(i \neq j \neq k = l) + \mathrm{perms} \to$ (1b) two sets of indices equal; three indices equal and one not; all indices equal.

- (1a) $\to (i; j \neq i; k \neq i \neq j; l \neq i \neq j \neq k)$. Averaging over the source galaxy positions gives:

$$\text{(1a)}_{\mathrm{part}} = P\left(\sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l Q_i Q_j Q_k Q_l \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k}\gamma_{\mathrm{t},l}\right)$$

$$= \prod_{\alpha=1}^{N} \left\{ \int \frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2} \right\} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l Q_i Q_j Q_k Q_l \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k}\gamma_{\mathrm{t},l}$$
$$= \frac{1}{(\pi\vartheta^2)^4} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l \int \mathrm{d}^2\theta_1 \ldots \mathrm{d}^2\theta_4 Q_1 \ldots Q_4 \gamma_{\mathrm{t},1} \ldots \gamma_{\mathrm{t},4}$$
$$= \frac{1}{(\pi\vartheta^2)^4} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l \mathcal{M}_{\mathrm{ap}}^4 \tag{2.74}$$

Finally, on averaging over the shear-field realisations and putting back the factors we see that:

$$\mathrm{E}\left[\text{(1a)}_{full}\right] = \frac{(\pi\vartheta^2)^4}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2} \frac{1}{(\pi\vartheta^2)^4} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l \left\langle \mathcal{M}_{\mathrm{ap}}^4 \right\rangle$$
$$= \frac{\sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2} \left\langle \mathcal{M}_{\mathrm{ap}}^4 \right\rangle \tag{2.75}$$

- (1b) $\rightarrow (i; j\neq i; k\neq i\neq j; l=i\neq j\neq k) + perms$. Averaging over the source galaxy positions gives:

$$\text{(1b)}_{part} = P\left( \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k Q_i^2 Q_j Q_k \gamma_{\mathrm{t},i}^2 \gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k} \right)$$
$$= \prod_{\alpha=1}^{N} \left\{ \int \frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2} \right\} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k Q_i^2 Q_j Q_k \gamma_{\mathrm{t},i}^2 \gamma_{\mathrm{t},j}\gamma_{\mathrm{t},k}$$
$$= \frac{1}{(\pi\vartheta^2)^3} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k \int \mathrm{d}^2\theta_1 \ldots \mathrm{d}^2\theta_3 Q_1^2 \ldots Q_3 \gamma_{\mathrm{t},1}^2 \ldots \gamma_{\mathrm{t},3}$$
$$= \frac{1}{(\pi\vartheta^2)^3} \sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k \mathcal{M}_{\mathrm{ap}}^2 \frac{\mathcal{M}_{\mathrm{s},2}}{\pi\vartheta^2} \,, \tag{2.76}$$

where in the above we have defined the quantity: $\mathcal{M}_{\mathrm{s},2} := \pi\vartheta^2 \int \mathrm{d}^2\theta_1 Q^2(\boldsymbol{\theta}_1)\gamma_{\mathrm{t}}^2(\boldsymbol{\theta}_1)$. We note that there are three identical contributions of this type that arise from the terms where $i=k$, $j=l$, and $j=k$. Hence, on averaging over the shear-field realisations and putting back the factors we see that the sum of these terms becomes:

$$\mathrm{E}\left[\text{(1b)}_{full}\right] = 4\frac{\sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2} \left\langle \mathcal{M}_{\mathrm{ap}}^2 \mathcal{M}_{\mathrm{s},2} \right\rangle \,. \tag{2.77}$$

- (1c) $\rightarrow (i; j\neq i; k=i; l=j) + (i; j\neq i; k=j; l=i)$. Averaging over the source galaxy positions gives:

$$\text{(1c)}_{part} = P\left( \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2 \gamma_{\mathrm{t},i}^2 \gamma_{\mathrm{t},j}^2 \right)$$
$$= \prod_{\alpha=1}^{N} \left\{ \int \frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2} \right\} \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2 \gamma_{\mathrm{t},i}^2 \gamma_{\mathrm{t},j}^2$$

$$= \frac{1}{(\pi\vartheta^2)^2} \sum_i \sum_{j\neq i} w_i^2 w_j^2 \int \mathrm{d}^2\theta_1 Q_1^2 \gamma_{\mathrm{t},1}^2 \int \mathrm{d}^2\theta_2 Q_2^2 \gamma_{\mathrm{t},2}^2$$
$$= \frac{1}{(\pi\vartheta^2)^2} \sum_i \sum_{j\neq i} w_i^2 w_j^2 \frac{\mathcal{M}_{\mathrm{s},2}^2}{(\pi\vartheta^2)^2} \,. \tag{2.78}$$

We note that the second term $(i; j \neq i; k = j; l = i)$ will be identical to the first after summation, and thus gives us an extra factor of 2. Hence, on averaging over the shear-field realisations and putting back the factors we see that the sum of these terms becomes:

$$\mathrm{E}\left[\text{(1c)}_{full}\right] = 2\frac{\sum_i w_i^2 \sum_{j\neq i} w_j^2}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2} \left\langle \mathcal{M}_{\mathrm{s},2}^2 \right\rangle \,. \tag{2.79}$$

The terms where three or four indices are the same vanish due to the constraints on the sums.

Turning now to the second term on the right-hand-side of Eq. (2.73). Performing the average over the source galaxy ellipticities we see that, since $i \neq j$ and $k \neq l$, we have two possibilities:

$$A\left[\epsilon_{\mathrm{t},i}^{(\mathrm{s})} \epsilon_{\mathrm{t},j}^{(\mathrm{s})} \epsilon_{\mathrm{t},k}^{(\mathrm{s})} \epsilon_{\mathrm{t},l}^{(\mathrm{s})}\right] = \frac{\sigma_\epsilon^4}{4}\left[\delta_{i,k}^K \delta_{j,l}^K + \delta_{i,l}^K \delta_{j,k}^K\right] \,, \tag{2.80}$$

Let us call these two terms (2a) and (2b). On performing the average over galaxy positions the first term becomes:

$$\text{(2a)}_{part} \equiv \mathrm{P}\left(\sum_i \sum_{j\neq i} \sum_k \sum_{k\neq l} w_i w_j w_k w_l Q_i Q_j Q_k Q_l \frac{\sigma_\epsilon^4}{4} \delta_{i,k}^K \delta_{j,l}^K\right)$$
$$= \mathrm{P}\left(\frac{\sigma_\epsilon^4}{4} \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2\right)$$
$$= \frac{\sigma_\epsilon^4}{4} \prod_{\alpha=1}^{N} \left\{\int \frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2}\right\} \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2$$
$$= \frac{\sigma_\epsilon^4}{4} \sum_i \sum_{j\neq i} w_i^2 w_j^2 \int \frac{\mathrm{d}^2\boldsymbol{\theta}_1}{\pi\vartheta^2} Q_1^2 \int \frac{\mathrm{d}^2\boldsymbol{\theta}_1}{\pi\vartheta^2} Q_2^2$$
$$= \frac{\sigma_\epsilon^4}{4} \sum_i \sum_{j\neq i} w_i^2 w_j^2 \frac{G^2}{(\pi\vartheta^2)^4} \,, \tag{2.81}$$

where in the above we have defined the quantity $G := \pi\vartheta^2 \int \mathrm{d}^2\theta_1 Q^2(\boldsymbol{\theta}_1)$. The second term (2b) is identical and so gives us a factor of 2. Finally the expectation over the ensemble of realisations of the shear field, with the normalisation factors restored, gives us:

$$\mathrm{E}\left[\text{(2a)}_{full} + \text{(2b)}_{full}\right] = \frac{\sum_i \sum_{j\neq i} w_i^2 w_j^2}{2\left(\sum_i \sum_{j\neq i} w_i w_j\right)^2} \sigma_\epsilon^4 G^2 \,. \tag{2.82}$$

Turning now to the third term on the right-hand-side of Eq. (2.73) we see that when summing over allowed indices the first and last terms in the square bracket will not contribute. Let us label

the remaining terms four terms (3a)–(3d). On averaging over the source galaxy positions the first of these terms can be written as:

$$\begin{aligned}
\text{(3a)}_{part} &= \mathrm{P}\left(\sum_i \sum_{j\neq i} \sum_k \sum_{l\neq k} w_i w_j w_k w_l Q_i Q_j Q_k Q_l \frac{\sigma_\epsilon^2}{2} \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},k}\delta^K_{jl}\right) \\
&= \frac{\sigma_\epsilon^2}{2}\mathrm{P}\left(\sum_i \sum_{j\neq i} \sum_{k\neq i\neq j} w_i w_j^2 w_k Q_i Q_j^2 Q_k \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},k} + \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2 \gamma^2_{\mathrm{t},i}\right) \\
&= \frac{\sigma_\epsilon^2}{2}\prod_{\alpha=1}^{N}\left\{\int \frac{\mathrm{d}^2\boldsymbol{\theta}_\alpha}{\pi\vartheta^2}\right\}\left(\sum_i \sum_{j\neq i} \sum_{k\neq i\neq j} w_i w_j^2 w_k Q_i Q_j^2 Q_k \gamma_{\mathrm{t},i}\gamma_{\mathrm{t},k}\right. \\
&\qquad\qquad \left. + \sum_i \sum_{j\neq i} w_i^2 w_j^2 Q_i^2 Q_j^2 \gamma^2_{\mathrm{t},i}\right) \qquad (2.83) \\
&= \frac{\sigma_\epsilon^2}{2}\sum_i \sum_{j\neq i} \sum_{k\neq i\neq j} w_i w_j^2 w_k \int \frac{\mathrm{d}^2\boldsymbol{\theta}_1}{\pi\vartheta^2} Q_1 \gamma_{\mathrm{t},1} \int \frac{\mathrm{d}^2\boldsymbol{\theta}_2}{\pi\vartheta^2} Q_2^2 \int \frac{\mathrm{d}^2\boldsymbol{\theta}_3}{\pi\vartheta^2} Q_3 \gamma_{\mathrm{t},2} \\
&+ \frac{\sigma_\epsilon^2}{2}\sum_i \sum_{j\neq i} w_i^2 w_j^2 \int \frac{\mathrm{d}^2\boldsymbol{\theta}_1}{\pi\vartheta^2} Q_1^2 \gamma^2_{\mathrm{t},1} \int \frac{\mathrm{d}^2\boldsymbol{\theta}_2}{\pi\vartheta^2} Q_2^2 \\
&= \frac{\sigma_\epsilon^2}{2(\pi\vartheta^2)^4}\sum_i \sum_{j\neq i} \sum_{k\neq i\neq j} w_i w_j^2 w_k \mathcal{M}^2_{\mathrm{ap}} G + \frac{\sigma_\epsilon^2}{2(\pi\vartheta^2)^4}\sum_i \sum_{j\neq i} w_i^2 w_j^2 \mathcal{M}_{\mathrm{s},2} G\,, \qquad (2.84)
\end{aligned}$$

If we repeat the above calculation for the other 3 remaining terms (3b)–(3d), then we see that they yield exactly the same answer. This means that when summing these contributions we simply multiply our answer by a factor of 4. Finally, the expectation over the ensemble of realisations of the shear field, with the normalisation factors restored, gives us:

$$\begin{aligned}
\mathrm{E}\left[\text{(3a)}_{full} + \cdots + \text{(3d)}_{full}\right] &= \frac{2\sum_i \sum_{j\neq i} \sum_{k\neq i\neq j} w_i w_j^2 w_k}{\left(\sum_i \sum_{j\neq i} w_i w_j\right)^2}\sigma_\epsilon^2 \mathcal{M}^2_{\mathrm{ap}} G \\
&\quad + \frac{2\sum_i \sum_{j\neq i} w_i^2 w_j^2}{\left(\sum_i \sum_{j\neq i} w_i w_j\right)^2}\sigma_\epsilon^2 \mathcal{M}_{\mathrm{s},2} G\,. \qquad (2.85)
\end{aligned}$$

On summing all of the contributions to the variance that come from terms Eq. (2.75), Eq. (2.77), Eq. (2.79), Eq. (2.82) and Eq. (2.85) we see that the weighted variance of the estimator for the aperture mass variance can be written:

$$\begin{aligned}
\mathrm{Var}\left[\widehat{M^2_{\mathrm{ap}}(\vartheta)}\right] &= \frac{\sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} \sum_{l\neq k\neq j\neq i} w_i w_j w_k w_l}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2}\left\langle \mathcal{M}^4_{\mathrm{ap}}\right\rangle \\
&\quad + 4\frac{\sum_i \sum_{j\neq i} \sum_{k\neq j\neq i} w_i^2 w_j w_k}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2}\left\langle \mathcal{M}^2_{\mathrm{ap}}\mathcal{M}_{\mathrm{s},2}\right\rangle \\
&\quad + 2\frac{\sum_i w_i^2 \sum_{j\neq i} w_j^2}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2}\left\langle \mathcal{M}^2_{\mathrm{s},2}\right\rangle + \frac{2\sum_i w_i \sum_{j\neq i} w_j^2 \sum_{k\neq i\neq j} w_k}{\left(\sum_i w_i \sum_{j\neq i} w_j\right)^2}\sigma_\epsilon^2 G\left\langle \mathcal{M}^2_{\mathrm{ap}}\right\rangle
\end{aligned}$$

$$+\frac{2\sum_i\sum_{j\neq i}w_i^2w_j^2}{\left(\sum_i\sum_{j\neq i}w_iw_j\right)^2}\sigma_\epsilon^2G\left\langle\mathcal{M}_s^2\right\rangle+\frac{\sum_i w_i^2\sum_{j\neq i}w_j^2}{2\left(\sum_i w_i\sum_{j\neq i}w_j\right)^2}\sigma_\epsilon^4G^2-\left\langle\mathcal{M}_{\rm ap}^2\right\rangle^2\ . \tag{2.86}$$

## 2.B Mask maps

In Figure 2.12 we show the survey masks for the W1, W2, W3 and W4 fields of the CFHTLenS. This figure clearly illustrates the problem with incomplete sky coverage due to the survey boundaries, the holes drilled due to the diffraction effects of bright stars and the gaps between chips. One can also notice that, while the W1, W2 and W4 fields are fairly flat projections, the W3 field clearly suffers more from the effects of the curved geometry of the sky.

## 2.C Maps for the W2-W4 CFHTLenS fields

In Figs 2.13 - 2.15 we show aperture maps for the remaining CFHTLenS fields. The row are identically structured as in Figs 2.3 - 2.4, but we only show the signal-to-noise map in the left column and the inverse variance weight map in the right column.

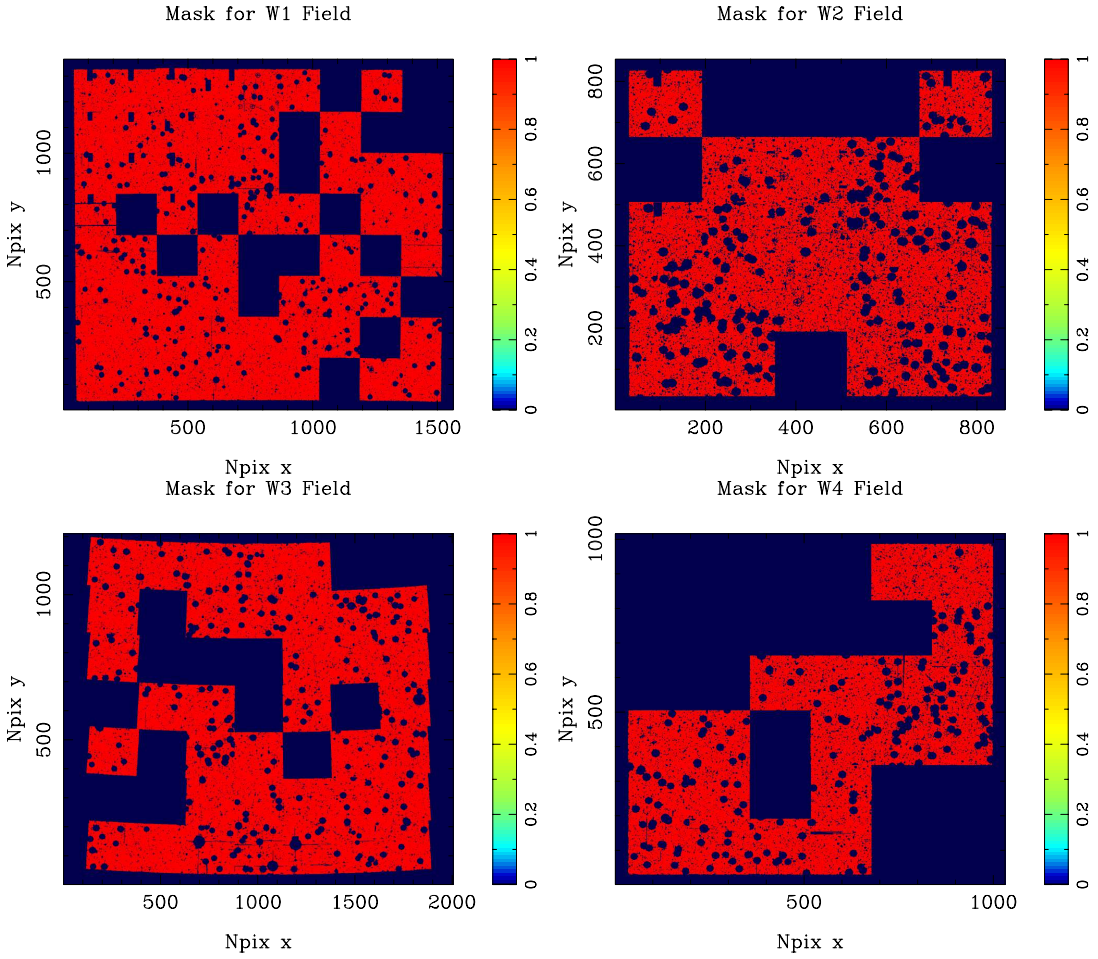


Figure 2.12 The pixel masks of the CFHTLenS. The top left, top right, bottom left and bottom right show the masks for the W1, W2, W3 and W4 fields, respectively. The red area shows the observed footprint.

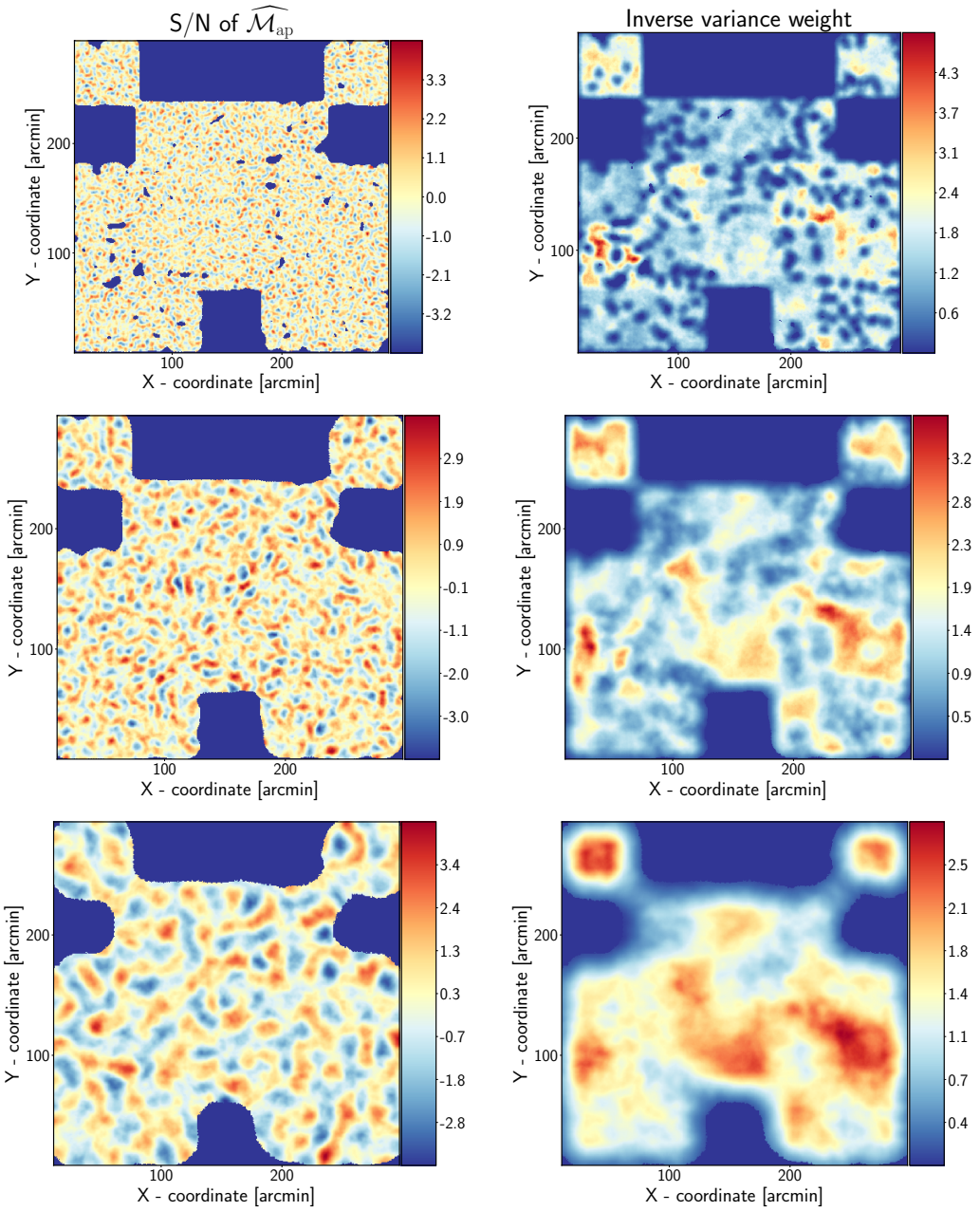


Figure 2.13 Signal-to-noise and aperture weight map of the W2 field of the CFHTLenS data.

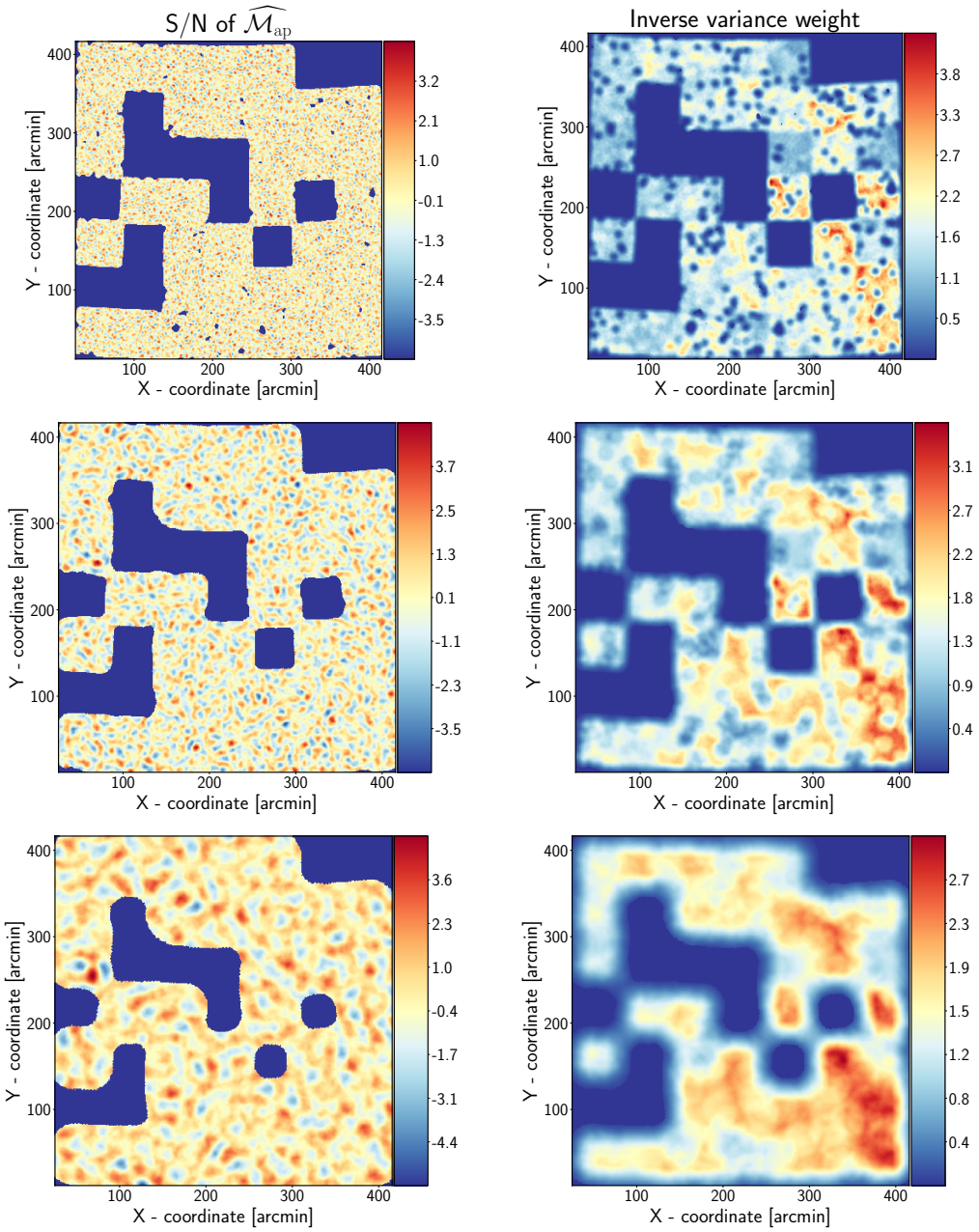


Figure 2.14 Signal-to-noise and aperture weight map of the W3 field of the CFHTLenS data.

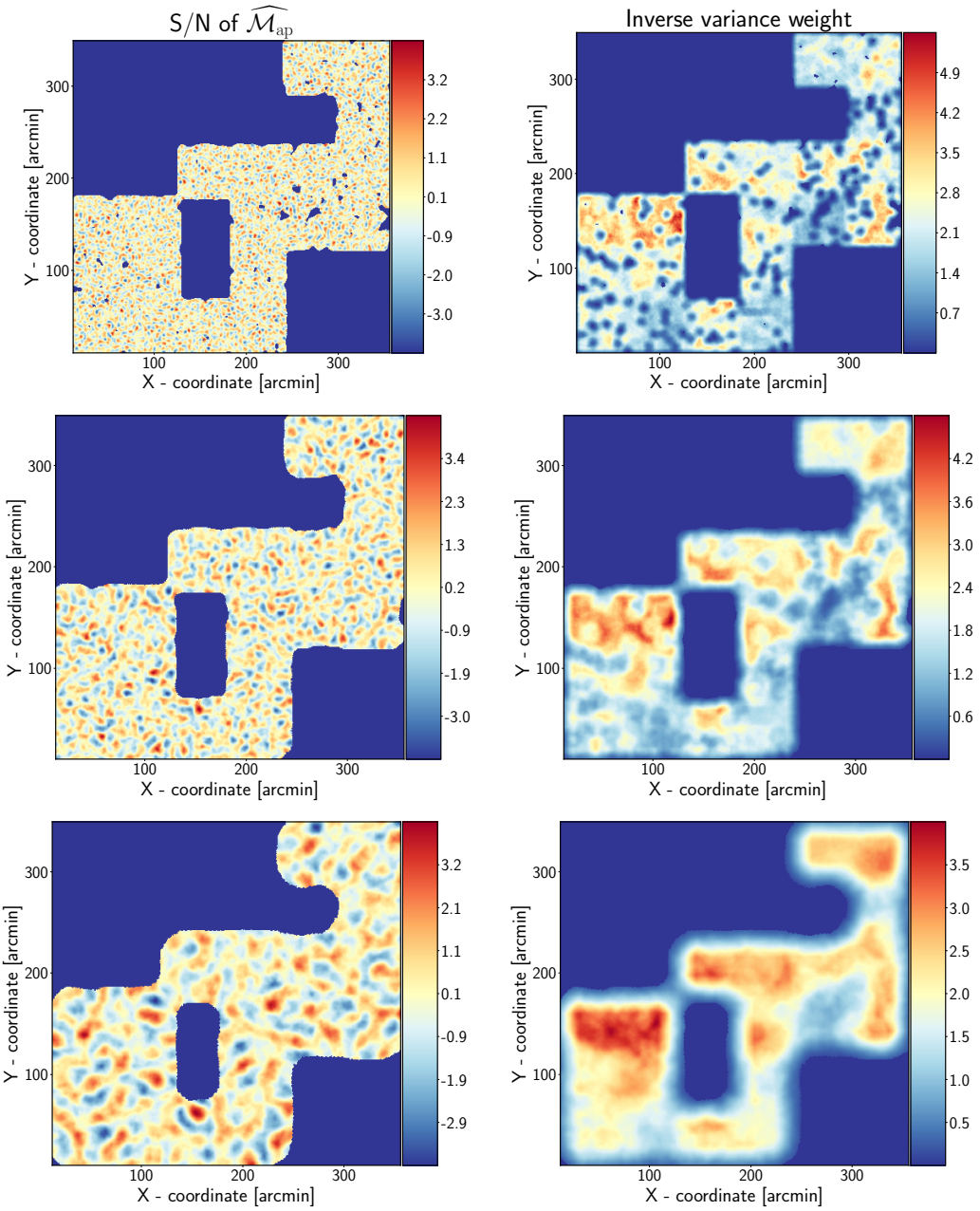


Figure 2.15 Signal-to-noise and aperture weight map of the W4 field of the CFHTLenS data.

# 3 Fast estimation of aperture mass statistics II

**Abstract**

We explore an alternative method to the usual shear correlation function approach for the estimation of aperture mass statistics in weak lensing survey data. Our approach builds on the direct estimator method. In this paper, we extend our analysis to statistics of arbitrary order and to the multiscale aperture mass statistics. We show that there always exists a linear order algorithm to retrieve any of these generalised aperture mass statistics from shape catalogs when the direct estimator approach is adopted. We validate our approach through application to a large number of Gaussian mock lensing surveys where the true answer is known and we do this up to 10th order statistics. We then apply our estimators to an ensemble of real-world mock catalogs obtained from $N$-body simulations – the SLICS mocks, and show that one can expect to retrieve detections of higher order clustering up to fifth order in a KiDS-1000 like survey. We expect that these methods will be of most utility for future wide-field surveys like Euclid and the Rubin Telescope.

## 3.1 Introduction

Weak gravitational lensing by large-scale structure of the light from distant galaxies is a powerful probe for constraining the cosmological parameters and distinguishing between competing models of the Universe (Blandford et al., 1991; Seitz et al., 1994; Jain & Seljak, 1997; Kaiser, 1998; Schneider et al., 1998; Zhang et al., 2007). The first measurements of the correlations in the shapes of distant background galaxies date back more than two decades (Bacon et al., 2000; Kaiser et al., 2000; Van Waerbeke et al., 2000; Wittman et al., 2000). Since then, cosmic shear observations have become ever more precise as the coupling of technological advancements and algorithmic developments have enabled us to conduct unprecedented deep optical imaging surveys of the

cosmos KiDS[1], DES[2] and HSC[3], with current state-of-the-art surveys now mapping thousands of square degrees (Hildebrandt et al., 2017; Troxel et al., 2018; Aihara et al., 2018; Hikage et al., 2019; Asgari et al., 2021). By the end of the decade planned experiments like *Euclid*[4] and the Rubin Telescope[5] (Laureijs et al., 2011; LSST, 2009) will map volumes close to the entire physical volume of our observable Universe. In order to make optimal use of these rich data sets we will need to push forward our understanding and modelling of various physical and measurement effects. In particular: accurate modelling of the nonlinear evolution of large-scale structure, including the baryonic physics effects; accurate modelling and correction of the point-spread function of the telescope; correcting the bias in the weak lensing shape estimation algorithms; and accounting for the intrinsic alignments, to name but a few of the main systematics (see Schneider, 2006b; Massey et al., 2013; Troxel & Ishak, 2015, for a more detailed discussion of these effects).

If the underlying matter density field were a Gaussian random field, then all of the information in a weak lensing survey would be contained in the shear two-point correlation function. However, physical effects like: the nonlinear growth of structure (Bernardeau et al., 2002), the mapping between cosmic shear and galaxy ellipticities (Miralda-Escude, 1991), and lensing beyond the Born approximation (Hilbert et al., 2009; Pratten & Lewis, 2016; Fabbian et al., 2018), all introduce non-Gaussianity in the maps. Furthermore, the nonlinear evolution also induces correlations in the convergence power spectrum multipoles, which grow stronger on small scales. This means that the information content of the second order statistics becomes saturated after a given multipole (Sato et al., 2011; Hilbert et al., 2012; Kayo et al., 2013; Marian et al., 2013; Byun et al., 2017). Thus in order to capture all of the cosmological information available in lensing surveys one must look to the higher order statistics of the shear field (Schneider et al., 1998; Bernardeau et al., 2002; Schneider & Lombardi, 2003). Furthermore, owing to the different ways in which the cosmological parameters and nuisance parameters depend on the higher-order statistics, the inclusion of such measurements brings with it the further virtue of being able to break parameter degeneracies, e.g. by combining second and third order statistics (Kilbinger & Schneider, 2005; Semboloni et al., 2011; Fu et al., 2014), or by incorporating the information found in the statistical properties of the peaks in the shear field (Marian et al., 2013; Kacprzak et al., 2016).

A powerful method to disentangle systematic effects from cosmic shear signals is the E/B decomposition (Crittenden et al., 2001; Schneider et al., 2002b). At leading order, pure weak

[1] kids.strw.leidenuniv.nl
[2] www.darkenergysurvey.org
[3] hsc.mtk.nao.ac.jp/ssp/
[4] www.cosmos.esa.int/web/euclid
[5] www.lsst.org

lensing signals are sourced by a scalar lensing potential, which means that their deflection fields are curl free. Equivalently, the ring-averaged cross component of the shear is expected to be zero (the B mode), while the tangential one contains all the lensing signal (the E mode). Thus B modes enable a robust test for the presence of systematic errors. One method to take advantage of this E/B decomposition is the so-called 'aperture mass statistics' (Kaiser, 1995; Schneider, 1996; Schneider et al., 1998). 'Aperture mass' ($\mathcal{M}_{\rm ap}$) and 'Map-Cross' ($\mathcal{M}_{\times}$) are obtained by convolving the tangential and cross shear with an isotropic filter function. Therefore by construction they are E/B-decomposed. Taking the second moment leads to the variance of aperture mass, the third to the skewness, the fourth to the kurtosis, etc.

The standard approach for measuring the aperture mass statistics in data utilises the fact that, for the flat sky, any $n$-point moment can be expressed in terms of integrals over the $n$-point shear correlation functions, modulo a kernel function (Schneider et al., 2002b; Jarvis et al., 2004). The reason for adopting this strategy stems from the fact that the correlation functions can reliably be estimated in the presence of a nontrivial survey mask. However, for these estimators to be accurate and E/B decomposed, one requires three conditions to be satisfied: (i) the $\xi_+/\xi_-$ correlations need to be measured down to zero separation; (ii) they also need to be measured up to a maximum angular scale, set by the exact form of the aperture mass filter and its angular scale; (iii) the angular bins must be sufficiently fine for the discretisation of the integrals to be reliable (Kilbinger & Schneider, 2005; Fu et al., 2014). Owing to galaxy image blending, signal-to-noise issues and the finite size of the survey, the lower bound is never possible and the upper bound means that biases can occur due to edge effects. In addition, while the mean estimate is unbiased, the variance matrix does require one to carefully account for the mask (Schneider et al., 2002a; Friedrich et al., 2016). More recent developments that also make use of the shear correlation functions, while circumventing the issues of E/B leakage on small scales are the ring statistics and COSEBIs (Schneider & Kilbinger, 2007b; Schneider et al., 2010). While those approaches can be in principle be extended to higher order statistics, the estimation of the $n$-point correlation functions turns out to be notoriously time consuming (Schneider et al., 2005; Jarvis et al., 2003). Further methods to extract non-Gaussian information from the aperture mass look at its probability density function as a whole (Bernardeau & Valageas, 2000; Munshi et al., 2004; Barthelemy et al., 2020) or at the distribution of its signal-to-noise peaks (Marian et al., 2012; Heydenreich et al., 2020; Martinet et al., 2021).

In Porth et al. (2020) we took a different approach and explored a computationally efficient (accelerated) implementation of the original direct estimator of the aperture mass dispersion (Schneider, 1998). Rather than measuring the correlation functions of the shear polar, in this

formulation one instead directly measures cumulants of $\mathcal{M}_{\rm ap}$ on a set of apertures and then uses an optimised weighting scheme to average the estimates, along with a restriction on the types of apertures that are acceptable. The present work extends our previous investigation in a number of important ways. First, we construct accelerated direct estimators for the higher-order aperture mass moments, including the skewness, kurtosis, etc. Second, we also develop further the multiscale aperture moments (Jarvis et al., 2003; Schneider et al., 2005). These two improvements enable us to better trace the full, harmonic mode, configuration dependence of the convergence polyspectra.

This paper is organised as follows: In §3.2 we introduce key concepts of weak lensing, define the aperture mass and show how its connected cumulants are related to the convergence polyspectra. In §3.3 we revisit the direct estimators for higher order aperture mass measures and construct suitable bases, in which each statistic can be computed in linear time complexity. After investigating the variance of the direct estimators, we give details of our updated algorithm used to perform the measurements. In §3.4 we empirically verify the linear scaling and the measurements of our implementation of the direct estimator on Gaussian mocks. In §3.5 we then apply the estimator to the SLICS simulation suite in order to assess up to which order one can expect to extract information from the aperture mass statistics on a KiDS-1000 like survey. Finally, in §3.6 we summarise our findings, conclude and discuss future work.

## 3.2 Higher order aperture mass measures for cosmic shear

### 3.2.1 Weak gravitational lensing and aperture mass

In this paper we are mainly concerned with the weak lensing of distant background (source) galaxy shapes by the intervening large-scale structure (for detailed reviews of the topic see Bartelmann & Schneider, 2001; Schneider, 2006a,b; Dodelson, 2003, 2017; Kilbinger, 2015; Mandelbaum, 2018). The two fundamental quantities describing this mapping from true to observed galaxy images are the convergence $\kappa$ and the complex shear $\gamma = \gamma_1 + i\gamma_2$, which, assuming a metric theory of gravity, are all derived from an underlying scalar lensing potential. In a galaxy survey the effective convergence at angular position $\boldsymbol{\theta}$ and radial comoving distance $\chi$ can be connected to the density contrast $\delta(\chi\boldsymbol{\theta}, \chi)$ through:

$$\kappa(\boldsymbol{\theta}) = \frac{3}{2}\Omega_{\rm m,0}\left(\frac{H_0}{c}\right)^2 \int_0^{\chi_{\rm H}} {\rm d}\chi' \frac{\chi'}{a(\chi')} g(\chi')\delta(\chi'\boldsymbol{\theta}, \chi') \,, \tag{3.1}$$

where $\Omega_{\rm m,0}$ is the total matter density, $H_0$ denotes the Hubble constant, $a$ is the scale factor, $c$ is the speed of light, $\chi_H$ is the comoving distance to the horizon and $g(\chi)$ is a weight function

related to the normalized redshift distribution $\mathrm{d}n(z)/\mathrm{d}z$ of the source galaxies as

$$g(\chi') \equiv \int_{z(\chi')}^{z_\mathrm{H}} \mathrm{d}z \, \frac{\mathrm{d}n(z)}{\mathrm{d}z} \frac{[\chi(z) - \chi']}{\chi(z)} \, . \tag{3.2}$$

Aperture mass was developed by Schneider (1996) as a technique to estimate projected mass overdensities enclosed within a circular region:

$$\mathcal{M}_\mathrm{ap}(\boldsymbol{\theta}_0; \vartheta) = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \kappa(\boldsymbol{\theta}_1) U(|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0|; \vartheta) \, , \tag{3.3}$$

where $U$ is a compensated filter function. In the flat sky limit the (cross) aperture mass can be expressed in terms of a related circularly symmetric filter function $Q(U)$ and the complex shear field $\gamma$ in its E/B-decomposed basis:

$$\mathcal{M}_{\mathrm{ap}/\times}(\boldsymbol{\theta}_0; \vartheta) \equiv \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}_1 \gamma_{\mathrm{t}/\times}(\boldsymbol{\theta}_1; \boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0|; \vartheta) \, , \tag{3.4}$$

where the tangential and cross components of the shear field at position $\boldsymbol{\theta} + \boldsymbol{\theta}_0$ with respect to the aperture center $\boldsymbol{\theta}_0$ are defined as (Bartelmann & Schneider, 2001):

$$\gamma_\mathrm{t}(\boldsymbol{\theta}; \boldsymbol{\theta}_0) + i\gamma_\times(\boldsymbol{\theta}; \boldsymbol{\theta}_0) \equiv -\gamma(\boldsymbol{\theta} + \boldsymbol{\theta}_0) \mathrm{e}^{-2i\phi} \, , \tag{3.5}$$

in which $\phi$ denotes the polar angle associated with the vector $\boldsymbol{\theta}$. In the absence of systematic errors (B-modes) in the lensing data, map-cross should vanish (Schneider et al., 2002b).

For this work we will make use of the polynomial filter function introduced by Schneider et al. (1998):

$$Q(\theta; \vartheta) = \frac{6}{\pi\vartheta^2} \left(\frac{\theta}{\vartheta}\right)^2 \left[1 - \left(\frac{\theta}{\vartheta}\right)^2\right] \mathcal{H}(\vartheta - \theta) \, , \tag{3.6}$$

where $\vartheta$ is the characteristic scale of the filter and $\mathcal{H}(x)$ is the Heaviside function, which guarantees that the filter function has compact support.

### 3.2.2 A hierarchy of aperture mass measures

One may construct moments of the aperture mass field, and this gives rise to the so called aperture mass statistics. At the two-point level this gives us the variance $\left\langle \mathcal{M}_\mathrm{ap}^2 \right\rangle_\mathrm{c}(\vartheta_1)$ and at the three-point, the skewness $\left\langle \mathcal{M}_\mathrm{ap}^3 \right\rangle_\mathrm{c}(\vartheta_1)$, etc., where the subscript c stands for the connected cumulant obtained from the moments (Scoccimarro & Frieman, 1996). Owing to the fact that the aperture mass is a convolution of the convergence field with a filter function, it is possible to rewrite these moments in terms of their Fourier space counterparts, that is the convergence spectra. For example for the variance and skewness we have:

$$\left\langle \mathcal{M}_\mathrm{ap}^2 \right\rangle_\mathrm{c}(\vartheta) = \int \frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2} C_{\kappa,2}(\boldsymbol{\ell}_1) \left|\widetilde{U}_\vartheta(\boldsymbol{\ell}_1)\right|^2 \, ; \tag{3.7}$$

$$\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm c}(\vartheta) = \int \frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2} \cdots \int \frac{{\rm d}^2\boldsymbol{\ell}_3}{(2\pi)^2}\, (2\pi)^2 \delta^D \left( \sum_{i=1}^{3} \boldsymbol{\ell}_i \right) \times C_\kappa(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_3)\, \widetilde{U}_\vartheta(\boldsymbol{\ell}_1) \widetilde{U}_\vartheta(\boldsymbol{\ell}_2) \widetilde{U}_\vartheta(\boldsymbol{\ell}_3) \ , \tag{3.8}$$

where $\widetilde{U}_{\vartheta_i}$ denotes the Fourier transform of the aperture mass filter function $U(\theta; \vartheta_i)$ and $C_\kappa(\boldsymbol{\ell}_1)$ denotes the convergence power spectrum, and $C_\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3)$ the convergence bispectrum. These spectra can formally be defined:

$$\langle \tilde{\kappa}(\boldsymbol{\ell}_1) \tilde{\kappa}(\boldsymbol{\ell}_2) \rangle_{\rm c} = (2\pi)^3 \delta^D (\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2)\ C_\kappa(\boldsymbol{\ell}_1) \ ; \tag{3.9}$$

$$\langle \tilde{\kappa}(\boldsymbol{\ell}_1) \tilde{\kappa}(\boldsymbol{\ell}_2) \tilde{\kappa}(\boldsymbol{\ell}_3) \rangle_{\rm c} = (2\pi)^3 \delta^D \left( \sum_{i=1}^{3} \boldsymbol{\ell}_i \right)\ C_\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) \ . \tag{3.10}$$

This of course can be generalised to $n$-point aperture mass moments:

$$\left\langle \mathcal{M}_{\rm ap}^n \right\rangle_{\rm c}(\vartheta) = \int \frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2} \cdots \int \frac{{\rm d}^2\boldsymbol{\ell}_n}{(2\pi)^2}\, (2\pi)^2 \delta^D \left( \sum_{i=1}^{n} \boldsymbol{\ell}_i \right) \times C_\kappa(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_n)\, \widetilde{U}_\vartheta(\boldsymbol{\ell}_1) \cdots \widetilde{U}_\vartheta(\boldsymbol{\ell}_n) \ , \tag{3.11}$$

where the $n$-point convergence spectrum is defined:

$$\langle \tilde{\kappa}(\boldsymbol{\ell}_1) \ldots \tilde{\kappa}(\boldsymbol{\ell}_n) \rangle_{\rm c} = (2\pi)^3 \delta^D \left( \sum_{i=1}^{n} \boldsymbol{\ell}_i \right)\ C_\kappa(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n) \ . \tag{3.12}$$

It is worth noting that due to the fact that $\widetilde{U}$ is a sharply peaked filter function in Fourier space, the aperture mass moment on a given scale only carries information about a specific range of wavemodes $\boldsymbol{\ell}$ from the underlying polyspectrum. In order to extract more of the information that is available one needs to compute Eq. (3.11) for a large set of aperture radii (Schneider et al., 2005).

### 3.2.3 Multiscale aperture mass moments and their correlators

Even if one considers a wide range of aperture radii there will be certain wavemode configurations of the polyspectra that are suppressed when compared with other configurations. This may result in a loss of sensitivity to certain physical effects that are only manifest in the higher-order polyspectra, such as those induced by modifications of gravity or primordial non-Gaussianities. In order to combat this one can further generalise the aperture mass moments in several ways. First, if we choose different scales for the aperture mass filter function, then we get the multiscale aperture mass moments. For the $n$-point multiscale aperture mass moment this can be written:

$$\left\langle \mathcal{M}_{\rm ap}^n \right\rangle_{\rm c}(\vartheta_1, \cdots, \vartheta_n) = \int \frac{{\rm d}^2\boldsymbol{\ell}_1}{(2\pi)^2} \cdots \int \frac{{\rm d}^2\boldsymbol{\ell}_n}{(2\pi)^2}\, (2\pi)^2 \delta^D \left( \sum_{i=1}^{n} \boldsymbol{\ell}_i \right)$$

$$\times\, C_\kappa(\boldsymbol{\ell}_1,\cdots,\boldsymbol{\ell}_n)\,\widetilde{U}_{\vartheta_1}(\boldsymbol{\ell}_1)\cdots\widetilde{U}_{\vartheta_n}(\boldsymbol{\ell}_n)\ . \tag{3.13}$$

Second, if we correlate a set of apertures at different spatial positions in the sky, then one can define the multiscale aperture mass moment correlators (Szapudi & Szalay, 1997; Munshi & Coles, 2003). There are two special cases where this approach can be applied, the first is the case where the separation of the aperture is directed perpendicular to the line of sight. The second case is where the apertures are placed along the same line of sight, but where different tomographic bins of source galaxies are used to estimate the aperture mass. The former case measures the correlation of the cumulants on the same redshift slice, but at different angular positions. The latter case corresponds to correlating aperture measures in different surveys with overlapping footprints, or between photometric redshift bins within the same survey. As the aperture mass filter carries most of its weight in a compact region surrounding the aperture center one expects the signal to fall off rapidly for aperture separations that exceed beyond a few times the aperture radius. Generalizing the result of (Schneider et al., 1998) we can formally write this as follows:

$$\begin{aligned}\left\langle \mathcal{M}_{\rm ap}^n \mathcal{M}_{\rm ap}^m \right\rangle_c (\vartheta_1,\cdots,\vartheta_n,\vartheta_1',\cdots,\vartheta_m';\overrightarrow{\Delta}) = \\ \int \frac{\mathrm{d}^2\boldsymbol{\ell}_1}{(2\pi)^2}\cdots\frac{\mathrm{d}^2\boldsymbol{\ell}_{n+m}}{(2\pi)^2}\,(2\pi)^2\delta^D\left(\sum_{j=1}^{n+m}\boldsymbol{\ell_j}\right)\,C_\kappa(\boldsymbol{\ell}_1,\cdots,\boldsymbol{\ell}_{n+m}) \\ \times\ \ \tilde{U}_{\vartheta_1}(\boldsymbol{\ell}_1)\cdots\tilde{U}_{\vartheta_m'}(\boldsymbol{\ell}_{n+m})\,e^{i\overrightarrow{\Delta}\sum_{j=1}^m \boldsymbol{\ell}_{n+j}}\ ,\end{aligned} \tag{3.14}$$

where $\boldsymbol{\Delta}$ is a separation vector. Note that for zero separation we recover the $(m+n)$th cumulant. In addition, we can assess the impact of the exponential factor by evaluating the two point cross-correlation coefficients $r_{mn}$, which are defined in a similar way to those in (Munshi & Valageas, 2005):

$$r_{mn}(\Delta) \equiv \frac{\langle X^m X^n\rangle_c(\Delta)}{\langle X^{m+n}\rangle_c}\ , \tag{3.15}$$

where for our case $X^m = \mathcal{M}_{\rm ap}^m$. In this work, however, we do not consider the cosmological information contained in Eq. (3.14), but instead use it to assess how fast the $r_{mn}$ converge to unity - this can be seen as a proxy for how densely apertures need to be sampled within a survey footprint to retrieve all available signal.

## 3.3 Estimators for higher order aperture mass statistics

### 3.3.1 Direct estimators for the aperture mass moments and their evaluation in linear order time

In this subsection we concern ourselves with estimators for higher order aperture mass statistics that mimic the original theoretical expressions Eq. (3.11) more closely. At first, let us investigate the special case of all the radii being equal.

Consider an aperture of angular radius $\vartheta$, centred on the position $\boldsymbol{\theta}_0$. The aperture contains $N$ galaxies[6] with positions $\boldsymbol{\theta}_i$, complex ellipticities $e_i$ and weights $w_i$. Then, for a single aperture, one can write down an estimator for the $n$th order aperture mass statistic Eq. (3.11) as (Schneider et al., 1998; Munshi & Coles, 2003)

$$\widehat{M_{\mathrm{ap}}^n} = (\pi\vartheta^2)^n \frac{\sum_{(i_1,\dots,i_n)^N} w_{i_1} Q_{i_1} e_{t,i_1} \cdots w_{i_n} Q_{i_n} e_{t,i_n}}{\sum_{(i_1,\dots,i_n)^N} w_{i_1} \cdots w_{i_n}} \,, \tag{3.16}$$

where we defined the shorthand notation[7]

$$\sum_{(i_1,\dots,i_n)^N} \equiv \sum_{i_1=1}^{N} \sum_{i_2\neq i_1}^{N} \cdots \sum_{i_n\neq i_{n-1}\neq\dots\neq i_1}^{N} \,. \tag{3.17}$$

On applying the above estimator to the case of $n = 2$, one can easily show that that this estimator is unbiased after averaging over the intrinsic ellipticity distribution, the galaxy positions within the aperture, and finally over cosmological ensembles (Schneider et al., 1998; Porth et al., 2020).

If we were to apply the above estimator given by Eq. (3.16) to determine the hierarchy of aperture mass moments, then this naive implementation would appear to result in an estimator that requires of the order $N^n$ operations to compute. However, following our earlier work (Porth et al., 2020), one can complete the sums to transform the estimators into sums and products of linear order terms. In Appendix 3.A we explicitly show, using elementary means, how one can compute the skewness ($\widehat{M_{\mathrm{ap}}^3}$) and kurtosis ($\widehat{M_{\mathrm{ap}}^4}$) using linear sums. The results for second, third and fourth orders are:

$$\widehat{M_{\mathrm{ap}}} = M_{\mathrm{s},1} \,; \tag{3.18}$$

$$\widehat{M_{\mathrm{ap}}^2} = \frac{M_{\mathrm{s},1}^2 - M_{\mathrm{s},2}}{1 - S_2} \,; \tag{3.19}$$

[6] Strictly speaking, we select galaxies within the support of the $Q$ filter function of that aperture. For the filter functions we use in this work the support is always concentric around the aperture center and linearly scaling with aperture radius. Therefore we will continue referring to $N$ as the number of galaxies per aperture.

[7] Note that in this definition and in general the notation $i_1 \neq i_2 \neq \cdots \neq i_n$ below a single summation symbol indicates that the summation index (which is understood to be the first one in the expression) is not equal to any of the other indices in the expression.

$$\widehat{M_{\rm ap}^3} = \frac{M_{\rm s,1}^3 - 3M_{\rm s,2}M_{\rm s,1} + 2M_{\rm s,3}}{1 - 3S_2 + 2S_3} \; ; \tag{3.20}$$

$$\widehat{M_{\rm ap}^4} = \frac{M_{\rm s,1}^4 - 6M_{\rm s,2}M_{\rm s,1}^2 + 3M_{\rm s,2}^2 + 8M_{\rm s,3}M_{\rm s,1} - 6M_{\rm s,4}}{1 - 6S_2 + 3\left(S_2\right)^2 + 8S_3 - 6S_4} \; , \tag{3.21}$$

where we have introduced two additional quantities: $S_m$ and $M_{{\rm s},m}$, which are defined:

$$M_{{\rm s},m} \equiv S_m \left(\pi\vartheta^2\right)^m \frac{\sum_{i=1}^N w_i^m Q_i^m e_{t,i}^m}{\sum_{i=1}^N w_i^m} \; ; \tag{3.22}$$

$$S_m \equiv \frac{\sum_{i=1}^N w_i^m}{\left(\sum_{i=1}^N w_i\right)^m} \; ; \tag{3.23}$$

Applying the elementary approach described in Appendix 3.A beyond fourth order rapidly becomes cumbersome, to say the least. We have therefore developed an analytic method for generation of the $n$th order estimator decomposed into linear sums. This follows from noting that the sum in Eq. (3.16) runs over unequal indices and that one can express any statistic $\widehat{M_{\rm ap}^n}$ as a sum of the power sums Eq. (3.22), where the coefficients preceding each term are determined with the help of the complete Bell polynomials $B_n$. Hence, for the general $n$th order estimate one has:

$$\widehat{M_{\rm ap}^n} = \frac{B_n\left(-M_{\rm s,1}, -M_{\rm s,2}, -2M_{\rm s,3}, ..., -(n-1)!M_{{\rm s},n}\right)}{B_n\left(-S_1, -S_2, -2S_3, ..., -(n-1)!S_n\right)} \; . \tag{3.24}$$

For full details of this derivation we refer the reader to Appendix 3.B. Here we only note that each argument that goes into Eq. (3.24) is a single sum over the galaxies in the aperture and is therefore independent of the order of the statistic. Using this formalism, we extend our decomposition to 5th and 6th order, giving us:

$$\begin{aligned}\widehat{M_{\rm ap}^5} = \frac{1}{N_5}\Big[&M_{\rm s,1}^5 - 10M_{\rm s,2}M_{\rm s,1}^3 + 15M_{\rm s,2}^2M_{\rm s,1} + 20M_{\rm s,3}M_{\rm s,1}^2 \\ &-20M_{\rm s,3}M_{\rm s,2} - 30M_{\rm s,4}M_{\rm s,1} + 24M_{\rm s,5}\Big] \; ;\end{aligned} \tag{3.25}$$

$$\begin{aligned}\widehat{M_{\rm ap}^6} = \frac{1}{N_6}\Big[&M_{\rm s,1}^6 - 15M_{\rm s,2}M_{\rm s,1}^4 + 45M_{\rm s,2}^2M_{\rm s,1}^2 - 15M_{\rm s,2}^3 + 40M_{\rm s,3}^2 \\ &- 90M_{\rm s,4}M_{\rm s,1}^2 + 40M_{\rm s,3}M_{\rm s,1}^3 - 120M_{\rm s,3}M_{\rm s,2}M_{\rm s,1} \\ &+90M_{\rm s,4}M_{\rm s,2} + 144M_{\rm s,5}M_{\rm s,1} - 120M_{\rm s,6}\Big] \quad ,\end{aligned} \tag{3.26}$$

where

$$N_5 = 1 - 10S_2 + 15\left(S_2\right)^2 + 20S_3 - 20S_3S_2 - 30S_4 + 24S_5 \; ; \tag{3.27}$$

$$\begin{aligned}N_6 =& 1 - 15S_2 + 45\left(S_2\right)^2 - 15\left(S_2\right)^3 + 40S_3 - 120S_3S_2 \\ &+ 40\left(S_3\right)^2 - 90S_4 + 90S_4S_2 + 144S_5 - 120S_6 \; .\end{aligned} \tag{3.28}$$

### 3.3.2 Direct estimators for the multiscale aperture mass moments

In complete analogy we can write down an unbiased direct estimator for the full multiscale aperture mass moments of Eq. (3.13):

$$\widehat{M_{\rm ap}^n}(\vartheta_1,...,\vartheta_n) = (\pi\vartheta_1^2)\cdots(\pi\vartheta_n^2) \times \frac{\sum_{(i_1,...,i_n)} w_{i_1} Q_{\vartheta_1,i_1} e_{t,i_1} \cdots w_{i_n} Q_{\vartheta_n,i_n} e_{t,i_n}}{\sum_{(i_1,...,i_n)^N} w_{i_1}\cdots w_{i_n}} \,, \tag{3.29}$$

where each index runs through all the galaxies within the aperture of the largest radius. In this case the power sums of Eqs. (3.22) and (3.23) do not form a sufficient basis to express these estimators, but we are still able to write down the estimators from elements within the sets

$$X_n \equiv \left\{ X^{(m)}_{(\alpha_1,...,\alpha_n)} \;\middle|\; \alpha_i \in \{0,1\}\,,\; \sum_{i=1}^n \alpha_i = m \le n \right\}\,, \tag{3.30}$$

where $X \in \{M_{\rm s}, S\}$ and the corresponding elements constitute of multivariate power sums being defined as

$$\begin{aligned} M^{(m)}_{{\rm s},\alpha_1,...,\alpha_n} &\equiv \left(\prod_{k=1}^n \left(\pi\vartheta_k^2\right)^{\alpha_k}\right) \sum_{i=1}^{N(\alpha)} w_i^m \prod_{j=1}^n \left[e_{t,i} Q_{\vartheta_j,i}\right]^{\alpha_j}\,, \\ S^{(m)}_{(\alpha_1,...,\alpha_n)} &\equiv \sum_{i=1}^{N(\alpha)} w_i^m\,, \end{aligned} \tag{3.31}$$

where $N(\alpha)$ denotes the number of galaxies within the aperture of the smallest radius for which $\alpha_i$ is not zero. Despite the more complicated looking form compared to the equal radius case these estimators can also be computed in $\mathcal{O}(N)$ time using the $\left|\hat{X}_n\right| = 2^n - 1$ distinct multivariate power sums Eq. (3.31) and summing over various partitions $P$ of the set $\{1,\cdots,n\}$:

$$\widehat{M_{\rm ap}^n}(\vartheta_1,...,\vartheta_n) = \frac{\sum_{m=1}^n \sum_{\pi\in P_{n,m}} (-1)^m \prod_{i=1}^m (n_i-1)!\; M^{(n_i)}_{{\rm s},\alpha_1(\pi_i),\cdots,\alpha_n(\pi_i)}}{\sum_{m=1}^n \sum_{\pi\in P_{n,m}} (-1)^m \prod_{i=1}^m (n_i-1)!\; S^{(n_i)}_{(\alpha_1(\pi_i),\cdots,\alpha_n(\pi_i))}}\,. \tag{3.32}$$

In this expression the combination of the two outer sums run through each partition $\pi$ that consists of $m$ blocks and the $\alpha(\pi_i)$ denote the value of the $\alpha$ as evaluated from the $i$th block of the partition. For a motivation of this equation and explicit expressions we again refer to Appendix 3.B.

### 3.3.3 Estimators applied to a large survey

In order to estimate any aperture statistics

$$\mathfrak{M} \in \left\{ \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle(\vartheta_1),\; \left\langle \mathcal{M}_{\rm ap}^3 \right\rangle(\vartheta_1), \ldots \right\} \tag{3.33}$$

on a contiguous survey field one can simply place an ensemble of apertures on the field and compute their weighted means

$$\hat{\mathfrak{M}} = \frac{\sum_{\rm ap} w_{\rm ap} \hat{\mathfrak{M}}_{\rm ap}}{\sum_{\rm ap} w_{\rm ap}} \,, \tag{3.34}$$

where the weights $w_i$ should be chosen to minimize the variance of the estimator. Owing to the linearity of Eq. (3.34), if the estimator of a single aperture is unbiased, then so is Eq. (3.34). Thus including more apertures will increase the signal-to-noise of the ensemble estimator.

### 3.3.4 Variance of the direct estimators

In order to understand how to weight the apertures we need to obtain expressions for the variance of the moment estimators. On generalizing the prescriptions outlined in (Schneider et al., 1998; Munshi & Coles, 2003) to include the shear weights, as in Porth et al. (2020), one can work out expressions for the variance of the higher-order direct estimators Eq. (3.16) for a given aperture. For the explicit derivation of the variance of the third order statistic see Appendix 3.D. From this analysis we see that the general expression can be written:

$$\begin{aligned}\sigma^2\left[\widehat{M^n_{\rm ap}}\right] = \sum_{\ell=0}^{n}\sum_{m=\ell}^{n} & \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_m}^2 w_{i_{m+1}} \cdots w_{i_n} w_{j_{m+1}} \cdots w_{j_n}}{\left(\sum_{\neq} w_{i_1} \cdots w_{i_n}\right)^2} \\ & \times C(n,\ell,m) M_{g,2}^{\ell} \left(\frac{\sigma_\epsilon^2}{2}\right)^{\ell} \langle \mathcal{M}_{\rm s,2}^{m-\ell} \mathcal{M}_{\rm ap}^{2(n-m)} \rangle \,,\end{aligned} \tag{3.35}$$

where the sum over the galaxy weights can again be decomposed as sums of (bivariate) power sums and the multiplicities are given by

$$C(n,\ell,m) = \frac{1}{\ell!(m-\ell)!}\left(\frac{n!}{(n-m)!}\right)^2 \,. \tag{3.36}$$

For a discussion of the origin for the multiplicity factor $C(n,\ell,m)$, as well as a motivation of Eq. (3.35) and some of its limits we refer the reader to Appendix 3.C; in particular we obtain for the shape noise dominated limit

$$\sigma^2\left[\widehat{M^n_{\rm ap}}\right] \approx n! \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_n}^2}{\left(\sum_{\neq} w_{i_1} \cdots w_{i_n}\right)^2} \left(\frac{\sigma_\epsilon^2}{2}\right)^n M_{g,2}^n \,. \tag{3.37}$$

The above formula gives the variance per aperture, thus for the estimator over the full survey field, Eq. (3.34), the variance can be written down as:

$$\begin{aligned}\sigma^2\left[\hat{\mathfrak{M}}\right] &= \mathrm{cov}\left(\frac{\sum_i w_i \hat{\mathfrak{M}}_i}{\sum_i w_i}, \frac{\sum_j w_j \hat{\mathfrak{M}}_j}{\sum_j w_j}\right) \\ &= S_2\, \sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right] + \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\rm ap}\right]}{\left(\sum_i w_i\right)^2} \sum_{i\neq j} w_i w_j \rho(\hat{\mathfrak{M}}_i, \hat{\mathfrak{M}}_j) \,,\end{aligned} \tag{3.38}$$

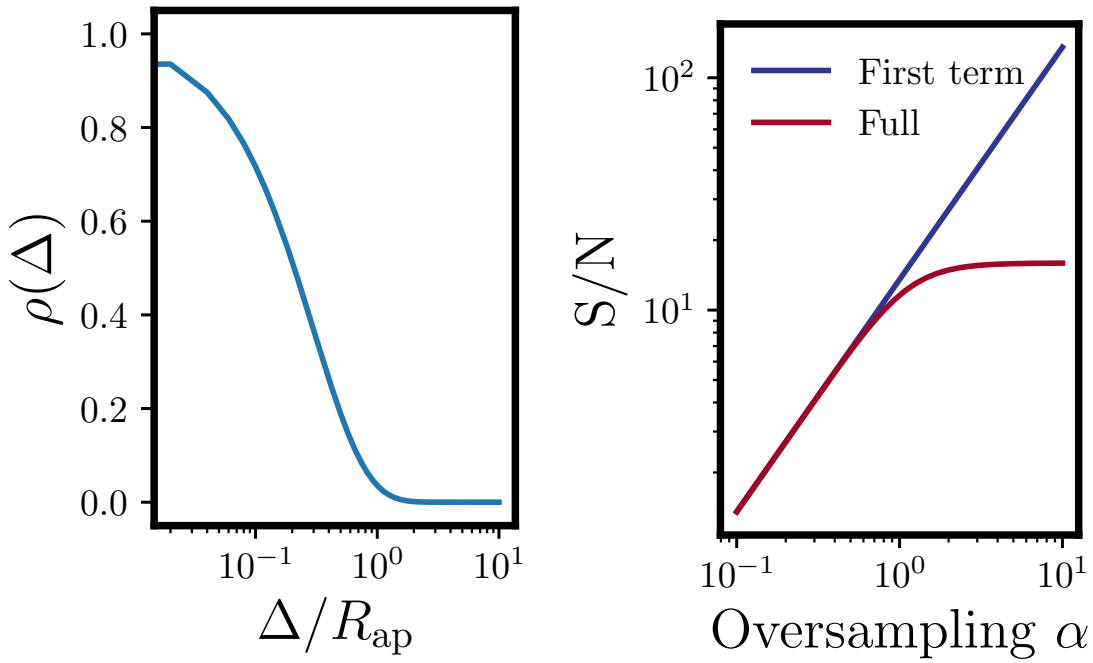


Figure 3.1 Example configuration of the correlation coefficient $\rho$ (left) and its effect on the signal contained in a survey field as predicted from Eq. (3.41) (right).

where in the above we have defined the cross-correlation coefficient between apertures whose centres are at position $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ to be:

$$\rho(\mathfrak{M}_i, \mathfrak{M}_j) \equiv \frac{\langle \mathfrak{M}_i, \mathfrak{M}_j \rangle}{\sqrt{\langle \mathfrak{M}_i, \mathfrak{M}_i \rangle \langle \mathfrak{M}_j, \mathfrak{M}_j \rangle}} \,. \tag{3.39}$$

Note that for the case of well separated apertures, the cross-correlation coefficient will vanish and only the first summand needs to be taken into account, which for unity weights gives the familiar $1/N_{\mathrm{ap}}$ scaling of the variance. If the apertures are oversampled, this assumption is no longer valid and the term involving $\rho$ must be included. Owing to the fact that $\rho$ should only depend on the relative spatial distance $\Delta$ between the aperture centres, we can rewrite (3.38) as a weighted sum over all possible distances between aperture center pairs:

$$\sigma^2\left[\hat{\mathfrak{M}}\right] = S_2\,\sigma^2\left[\hat{\mathfrak{M}}_{\mathrm{ap}}\right] + \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\mathrm{ap}}\right]}{\left(\sum_i w_i\right)^2} \sum_{b\in bins} \left( \sum_{i,j\in\mathcal{I}_b} w_i w_j \right) \rho(\hat{\mathfrak{M}}, \Delta_b) \tag{3.40}$$

$$\approx \frac{\sigma^2\left[\hat{\mathfrak{M}}_{\mathrm{ap}}\right]}{N_{\mathrm{ap}}} + 2\pi\,\frac{\sigma^2\left[\hat{\mathfrak{M}}_{\mathrm{ap}}\right]}{A_{\mathrm{survey}}} \int_{R_{\mathrm{ap}}/\alpha}^{\infty} \mathrm{d}\Delta\;\Delta\;\rho(\hat{\mathfrak{M}}, \Delta)\;, \tag{3.41}$$

where in the first step the bins are defined as a partition of the reals, and $\mathcal{I}_b \equiv \{i, j | \Delta(i, j) \in b\}$ collects all the aperture center pairs falling into bin $b$. For the second step we make the approximation that each aperture contains roughly the same signal such that the weights can be set to unity and we furthermore rewrote the expression in a continuous version, which makes

the interpretation of the cross term more concise. In particular, we parametrize the lower bound of the integral in terms of the aperture oversampling rate $\alpha \equiv R_{\rm ap}/\Delta_{\rm min}$.

In a realistic scenario we expect $\rho$ to to rapidly decrease from unity and then to approach zero for $\Delta \gg R_{\rm ap}$. An example of such a correlation coefficient is shown in Fig. 3.1. Here we explicitly see the importance of taking into account the cross term once there is a substantial overlap between neighbouring apertures. In this example we would infer that measuring the statistics with $\alpha \approx 2$ would be sufficient to extract most of the signal.

### 3.3.5 Implementation and scaling of the direct estimator

A practical implementation of Eq. (3.34) consists of three steps:

(i) Spatially organise the shape catalog to allow for a fast assignment of galaxies to apertures.

(ii) For each aperture of the ensemble compute $\hat{\mathfrak{M}}_{\rm ap}$, the associated weight and optionally additional systematics (i.e. the coverage fraction $c_k$). Store each of these values in an array.

(iii) Based on some aperture selection and aperture weighting criteria $w_{\rm ap}$, update the weights and evaluate the weighted sum.

In what follows we will explore each of these steps in more detail and for clarity we will denote the number of galaxies in the survey and in the aperture as $N_{\rm g}$ and $N_{\rm g,ap}$, respectively.

#### Assigning galaxies to apertures

For our implementation we use a spatial hashing data structure. We start by covering the survey footprint with an equal area mesh of $N_{\rm pix}$ pixels and create a hash table with the ID of each pixel as the key and the galaxy IDs as values. The hash function in our case is the ordinary pixel assignment function. For each aperture we iterate over the associated galaxies within pixels that partially lie within the $Q$ filter's support. The construction of the hash table scales as $\mathcal{O}\left(\max(N_{\rm pix}, N_{\rm g})\right)$ and the assignment is achieved in $\mathcal{O}\left(\max(N_{\rm pix,ap]}, N_{\rm g,ap})}\right)$ time per aperture. We found that when making a sensible choice of the mesh's coarseness, this data structure is more stable than a naive KD-tree based implementation as it does not require an additional range search operation which scales as $\mathcal{O}(\log(N_{\rm g}))$ per aperture and thus becomes a bottleneck for small apertures.

### Computing the statistics per aperture

For the case of all radii being equal we first compute the power sums in Eqs. (3.22) and (3.23) and then recursively transform them to the corresponding moments via the recurrence relation (Comtet, 1974)

$$B_{n+1}(x_1, \cdots, x_{n+1}) = \sum_{i=0}^{n} \binom{n}{i} B_n(x_1, \cdots, x_n) x_{i+1} \,, \tag{3.42}$$

where $B_0 \equiv 1$. Evaluating each power sum is linear in $N_{\mathrm{g,ap}}$ and for all practical applications the time taken for transforming to the $\mathcal{M}_{\mathrm{ap}}^n$ basis can be neglected.

For the general case we need to compute the relevant multivariable power sums Eq. (3.31) and bring them to the aperture moments basis by the transformation Eq. (3.32). In order to dynamically allocate and evaluate those expressions we use a combinadic counting scheme to organize the power sum basis whereas the transformation equation is generated with the help of restricted growth strings (Knuth, 2005).

### Choice of weights for the averaging

Following our findings in Porth et al. (2020) we employ an inverse shot noise weighting scheme with an additional hard cutoff $c_{\mathrm{lim}}$ for the aperture coverage $c_{\mathrm{ap}}$, which for second order statistics was found to lower the mask induced bias while increasing the signal-to-noise compared to equal weights. The explicit form of the weights for the $n$th moment can be found from Eq. (3.37) when neglecting all constant contributions:

$$\mathcal{W}_{\mathrm{ap}}^{(\mathrm{shot})}(c_{\mathrm{lim}}) \equiv \left[ \frac{\sum_{(i_1,\cdots,i_n)} w_{i_1}^2 \cdots w_{i_n}^2}{\left(\sum_{(i_1,\cdots,i_n)} w_{i_1} \cdots w_{i_n}\right)^2} \right]^{-1} \mathcal{H}(c_{\mathrm{ap}} - c_{\mathrm{lim}}) \,. \tag{3.43}$$

Dependent on whether we are dealing with the case of equal or unequal aperture radii the sums can be decomposed in a similar fashion as described above and evaluated together with the corresponding linearised direct estimator. As a further refinement one could also include the weights and completenesses of the surrounding apertures weighted by the spatial cross correlation coefficient $\hat{\rho}$ - this would upweight apertures that are close to a mask as they cover more unique area.

## 3.4 Results: application to Gaussian mocks

### 3.4.1 Aperture mass statistics and Gaussian lensing fields

In order to validate that our hierarchy of aperture mass moment estimators are unbiased and do indeed recover correct results, we first apply them to a set of Gaussian mock lensing simulations. In this case, the whole moment hierarchy can be written as powers of the variance of the aperture mass. Hence, this motivates us to define the scaled aperture mass moments:

$$s_n(\vartheta_1) \equiv \frac{1}{(n-1)!!} \frac{\left\langle \mathcal{M}_{\rm ap}^n \right\rangle (\vartheta_1)}{\left[ \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\vartheta_1) \right]^{n/2}} = \delta^K_{n,2\mathbb{N}} \,, \tag{3.44}$$

where the final equality is true for a Gaussian field only.

In order to test this we generated 256 Gaussian lensing mocks. The methodology to create each mock was as follows:

(i) We first generate a Gaussian convergence field over a $12 \times 12$ degrees$^2$ survey area. The area is tiled by a rectangular mesh of $8192^2$ pixels. The variance of the convergence is obtained through specifying the convergence power spectrum, and we do this for a source distribution similar to that for the CFHTLenS survey (Fu et al., 2014).

(ii) We next obtain the shear field. This is done by Fourier transforming the convergence field and making use of the Kaiser & Squires (1993) approach[8].

(iii) We then sample $4 \times 10^6$ galaxies into the survey footprint and use a multilinear interpolation of the shear field onto each galaxy.

Note that since we are assessing the accuracy of the estimators only we choose to set the intrinsic ellipticities of our source galaxies to zero. On repeating the analysis below when including this term we did not find a shift of the curves.

### 3.4.2 Computational scaling tests

Owing to the fact that each step of our algorithm is strictly linear, we expect a linear relationship between the elapsed time for estimator evaluation and the number of galaxies, for any given statistic. In addition, for the equal radius case the order of the statistics should not strongly

[8] In order to suppress edge effects introduced by the FFT we build the pixelated convergence field on an a plane having 16 times the area of the mock.

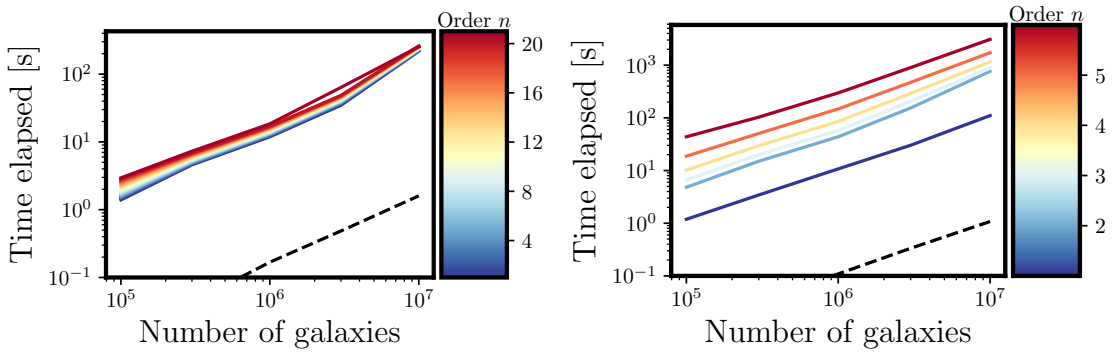


Figure 3.2 Computational complexity of the direct estimators for equal (left) and unequal (right) aperture radii as a function of the number of galaxies. All results are given for apertures of radius $10'$ which are oversampled by a factor of sixteen ($\alpha = 4$) on a survey field of size $(12\text{deg})^2$. Different colors indicate different orders of the evaluated statistics. The black dashed line indicates the time spent in constructing the spatial hash. We see that for equal aperture radii the evaluation of higher order statistics basically comes for free, while for unequal radii there is a constant multiplicative offset based on the relative size of the radii and on the order which traces the number of multivariate power sums that need to be evaluated. All the scaling were obtained when running the estimator on a single CPU core.

impact the evaluation time. However, for the unequal radius case, this does not necessarily hold true, since the computation depends on the relative sizes of the apertures as well as on the order of the statistics to be evaluated.

Figure 3.2 shows the elapsed time of the direct estimator calculation for a Gaussian mock, where the number of sampled galaxies in the mock is increased. Focusing on the left panel first, this shows the case for the standard aperture mass estimators with equal radii and here we compute all of the moments up to the 20th order. As expected the computational time for all of the moments scales linearly with the size of the problem and we also see that there is no obvious drop in performance for the higher order moments.

The right panel of Figure 3.2 is the same as the left panel, but now for the case of unequal radii aperture mass moments, and here we only consider moments up to 6th order. There are two differences between the equal and non-equal radius case. First, we can see that there is a much larger multiplicative offset between adjacent orders for the generalized statistics. This is expected as the number of basis elements that need to be allocated in that case is given by $2^n$ compared to the $n$ ones in the equal radius statistics. We also observe that for the second order statistic the unequal radius calculation does roughly need four times as long as the equal radius one. We can explain this offset when noting that for our example the ratio of the largest and

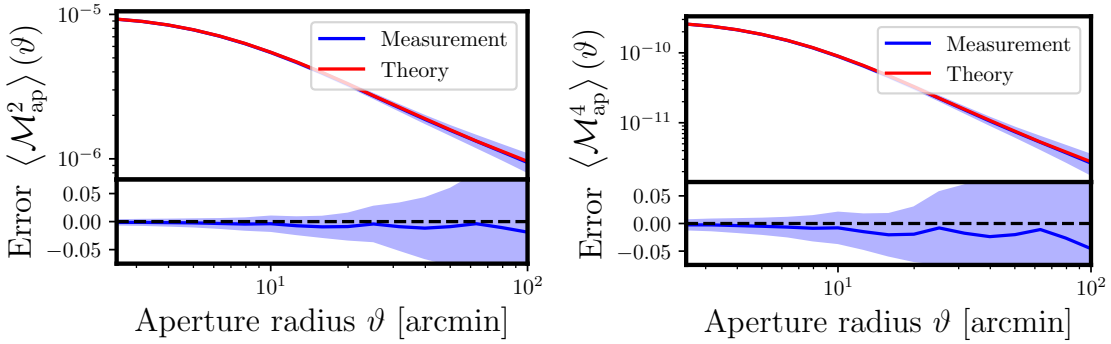


Figure 3.3 Comparison of the measured aperture mass moments with their theoretical prediction. *Left hand side:* The upper subpanel shows the aperture mass dispersion as a function of the aperture scale. The red line shows the theoretical predictions evaluated from the input power spectrum and the blue line shows the measurement from the mocks. The blue shaded regions show the standard deviation of the corresponding measurement across the ensemble. The lower subpanel shows the relative error between the measurement and the theory, with the line styles as before. *Right hand side:* Same as left panel, but for the kurtosis of aperture mass.

smallest scale was set to two. With our definition of the oversampling rate as being relative to the smallest aperture radius this implies that we need to allocate four times as many galaxies.

Finally, we note that the superior scaling of the direct estimator compared to traditional estimation methods should not come as a surprise. Looking back at the original definition Eq. (3.11) of the aperture mass one sees that it depends on the positions and shapes of the galaxies with respect to the aperture origin. In contrast, when switching to the description of aperture mass in terms of the shear correlation functions (i.e. Schneider et al., 2002b), the main dependence shifts to the relative distance and shapes between tuples of galaxies. This change of reference position makes the evaluation of correlation function based estimators intrinsically much more complex than a simple discretization of Eq. (3.11).

### 3.4.3 A hierarchy of aperture mass moments

Figure 3.3 shows a comparison of the direct estimators for the second and fourth order aperture mass moments as a function of angular scale as applied to the 256 Gaussian mocks. Here we consider the case where all the aperture radii are equal (recall that for a Gaussian field all of the odd moments vanish). In both cases the curves are in very good agreement with the Gaussian theory predictions, indicated by the solid red lines. We also note that for increasingly large aperture radii the measured results appear to be slightly below the theoretical expectation. This

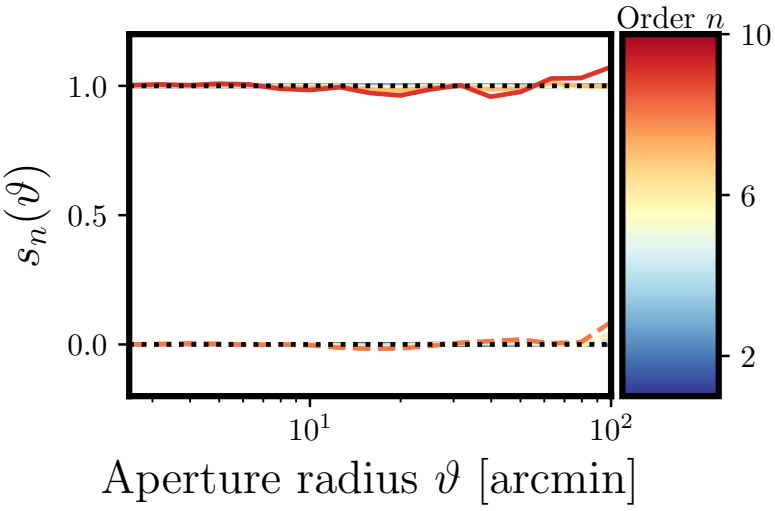


Figure 3.4 Scaled $n$th order aperture mass moments $s_n(\vartheta)$ (see Eq. 3.44), measured in the ensemble of 256 Gaussian mocks, as a function of the aperture scale, for all moments up to 10th order. The solid lines of varying colours show the mean of the measurements. The dotted black lines show the Gaussian theoretical expectations. For a Gaussian mock, the even order $s_n(\vartheta)$ give unity, and the odd ones vanish.

discrepancy can be attributed to finite field effects, as well as border effects being introduced by the Kaiser-Squires inversion method (see Pires, S. et al., 2020, for a discussion).

Figure 3.4 presents the measured $s_n$ (see Eq. (3.44)) for all of the aperture mass moments up to 10th order as a function of the aperture scale. We see that they are consistent with the Gaussian theoretical expectations. Note that in order to obtain this good agreement and circumvent the finite field effects described above, we used the ensemble mean of the measured aperture mass variance as the denominator in $s_n$.

Figure 3.5 displays the fourth and sixth order multiscale aperture mass statistics as a function of the scale parameter. Note that there are a number of options for exploring the configuration dependence of the multiscale aperture mass moments, here we focus on fixing the ratio of the filter lengths and varying the overall scale of the configuration with the parameter $\vartheta$, e.g. for the kurtosis we would have

$$\left\langle \mathcal{M}_{\rm ap}^n \right\rangle_{\rm c} (\vartheta_1, \cdots, \vartheta_n) = \left\langle \mathcal{M}_{\rm ap}^n \right\rangle_{\rm c} (\alpha_1 \vartheta, \cdots, \alpha_n \vartheta) \tag{3.45}$$

with $\alpha_i = a_i$, where $a_i$ is a constant vector (it is to be understood that $\alpha_i$ is different from the aperture spacing parameter $\alpha$). The estimates shown in the figure were obtained using our generalized estimator Eq. (3.32). As for the previous cases, we find good agreement between the measurements and the Gaussian predictions, which were obtained by making use of Eq. (3.13)

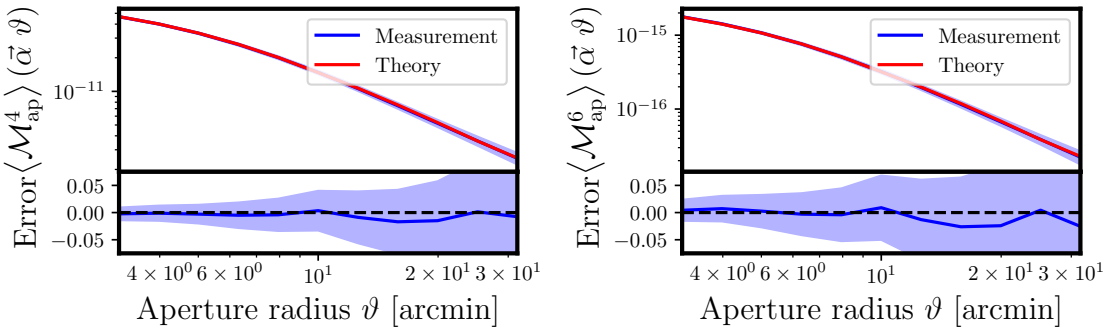


Figure 3.5 Multiscale aperture mass moments as a function of the scale parameter $\vartheta$, measured in the ensemble of 256 Gaussian mocks. Line styles are the same as in Fig. 3.3. *Left hand side:* the fourth order aperture statistics. In this case, the vector of aperture scales was set to $\boldsymbol{\alpha} = (0.5, 0.8, 1., 2.)$. *Right hand side:* Same as left hand side, but this time for the sixth order statistic with the vector of aperture scales set to $\boldsymbol{\alpha} = (0.5, 0.7, 1., 1.1, 1.5, 2.)$.

and Wicks theorem for the convergence polyspectra (Bernardeau et al., 2002).

## 3.5 Results: Detection significance of higher order moments

In this section we now turn to the question of the detection significance of higher order aperture statistics from current and future surveys.

### 3.5.1 The SLICS mocks

In order to answer this question we make use of the SLICS[9] mocks – this is a large suite of lensing mock catalogues generated from a large set of cosmological $N$-body simulations (for full details see Harnois-Déraps et al., 2018). Each SLICS mock corresponds to a survey area of $100\,\mathrm{deg}^2$. These are generated from the past light cone extracted from fully independent gravity-only $N$-body simulations, which evolve $N = 1532^3$ particles within a comoving box of length $L = 505\ h^{-1}\mathrm{M}pc$. The lensing maps are constructed using the Born approximation. We adopt the catalogues for which the galaxies are randomly distributed within the lightcone according to the KiDS-450 source distribution (Hildebrandt et al., 2017). The shape noise has been set to $\sigma_\epsilon = 0.29$ per shear component. In order to mimic a constraining power that is comparable to the KiDS-1000 data while not being too noisy, we rescale the errorbars by a factor of $\sqrt{10}$. This provides us with effectively 819 simulated $1000(\mathrm{deg})^2$ surveys with which to perform our analysis.

[9] https://slics.roe.ac.uk/

When estimating the aperture mass statistics from the SLICS mocks using the estimator given by Eq. (3.34), the achievable signal-to-noise ratio will depend on the number of sampled apertures selected. If too few are chosen then our estimate will be inefficient, on the other hand due to the fact that there are aperture-to-aperture correlations choosing too many will capture all of the available information, but ultimately will be computationally inefficient. We therefore expect that the information will saturate for a given oversampling rate, and that to sample at a higher rate would be of little use. To investigate this we proceed as in Porth et al. (2020) and place apertures on a regular grid with spacing $\Delta$, corresponding to an aperture oversampling rate of $\alpha \equiv \min(\{\vartheta_i\})/2\Delta$.

### 3.5.2 Measurement in the SLICS mocks

Figure 3.6 shows the detection significance of the equal radii aperture mass statistics for the second, third, fourth and fifth order aperture mass statistics as a function of the aperture scale and for various choices of the oversampling rate. For the second order statistics we also plot the theoretical prediction of the aperture mass dispersion evaluated from Eq. (3.11), where the convergence power spectrum was computed with CCL[10] (Chisari et al., 2019) using Halofit (Smith et al., 2003), but with the modifications of Takahashi et al. (2012), as the matter power spectrum. While part of the difference between the curves for small aperture radii could be attributed to uncertainties in Halofit, our suspicion is that they mostly stem from the limited particle mass resolution in the SLICS mocks (see Fig. 6 in Harnois-Déraps et al. (2018) for the resulting suppression of the shear correlation functions for small separations).

Several important points are worth noting from these measurements. First, we see that for a KiDS-1000 like survey there is sufficient fidelity to detect the aperture mass statistics up to fifth order, with the signal-to-noise peaking at an aperture size of around $\vartheta \approx 10'$ for all statistics. This is exciting, as this has never before been achieved with standard correlation function based estimators, and if correct would represent the first robust detection of these statistics using these methods. Second, while for the case of the two-point statistics the signal-to-noise ratio (shown in lower sub-panels for each plot) falls off slowly for larger apertures, this ratio approaches zero faster for the connected parts of the higher order statistics[11]. Third, while an aperture oversampling rate of $\alpha \approx 2$ seems sufficient to capture all the signal for second order statistics, it misses some information for subsequent orders where it becomes necessary to use $\alpha \gtrsim 4$.

[10] https://github.com/LSSTDESC/CCL

[11] Owing to the fact that the aperture mass has zero mean, the full and connected moments differ only for even order moments of four or more.

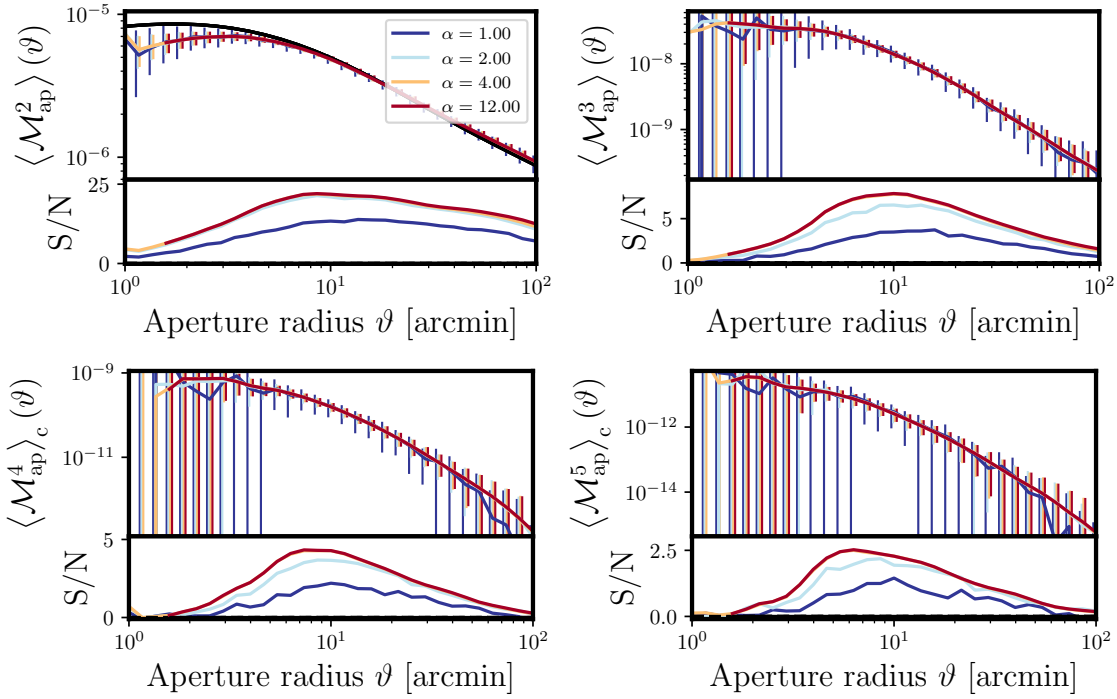


Figure 3.6 Measurement of the aperture mass statistics in the SLICS simulation suite for different aperture oversampling rates $\alpha$. All measurements were done on an ensemble of 819 realizations with an angular area of 100 deg$^2$ each, where the $n(z)$ follows the KiDS-450 distribution. The upper part of the panels correspond to the mean and rescaled standard deviation from the ensembles. The lower panel shows the signal-to-noise for the corresponding statistics when rescaled to match a 1000 deg$^2$ survey. For the aperture mass dispersion we additionally plot the theoretical prediction as the black line. For the fourth and fifth order plots we restrict ourselves to the contribution of the connected part of the convergence polyspectra. We see that choosing an oversampling parameter of $\alpha \gtrsim 4$ recovers most of the information.

Figure 3.7 displays the correlation structure of the aperture mass cumulants as well as the cumulative detection significance. We only consider measurements with $\vartheta \geq 10'$ as this is where the SLICS mocks do agree reasonably well with the theoretical predictions and and due to the fact that the robust theoretical modelling of those statistics might reach its limits at around those scales. We see that while for shape noise free ellipticity catalogs there are strong correlations for small aperture radii, this is not the case for the realistic mocks in which those scales are still shape noise dominated. We further note large correlations around the diagonal between different orders, where the degree of correlation increases with the order of the cumulants. For the cumulative detection significance we see that that the cumulants beyond third order do not add a substantial amplitude to the cumulative signal-to-noise. This is expected, given the relatively lower signal-to-noise as well as the larger portion of cross-covariances that need to be taken into

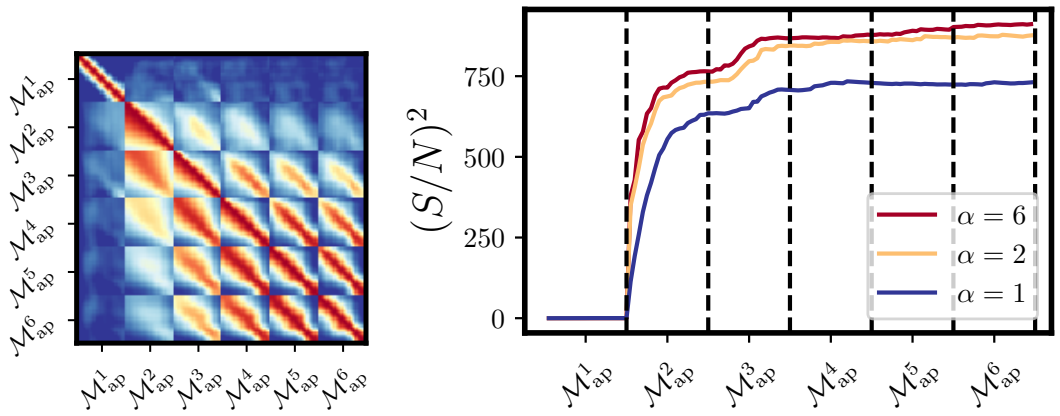


Figure 3.7 Correlation coefficient matrix (left) and cumulative detection significance (right) for connected moments of the aperture mass statistics. We take into account aperture sizes between $10'$ and $100'$. In the correlation matrix the lower triangle shows the results without shape noise while the upper part includes this term and serves as the basis for the computation of the detection significance.

account. One should note that this type of analysis does not imply that the higher order cumulants are obsolete as they still may add complementary information by breaking cosmological parameter degeneracies.

### 3.5.3 Multiscale aperture mass measurements

We now shift to the measurement of the multiscale statistics for which there are a number of ways on how to select various aperture scale multiplets. In Figure 3.8 we focus on a fixed set of aperture propositions and then simply scale them with a single parameter $\vartheta$. The different configurations of aperture radii that we have employed are shown in Table 3.1. We see that for both, the third and the fifth order moments there does not appear to be a strong decline in detection significance for multiscale apertures compared to the associated moments, even if the relative spread of radii is large.

Another way to select aperture scale multiplets for a statistic of order $n$ is choose a list of $m \geq n$ aperture scales and to compute the statistics for each choice of $n$ elements within that list. For our purposes we choose the subset in which none of the aperture radii are equal, as this speeds up our calculation, see Appendix 3.B.3 for the details. In the left hand side of Fig. 3.9 we show our measurements for the second, third and fourth order connected cumulants of the multiscale aperture mass statistic using ten logarithmically spaced scales between $5'$ and $50'$. The first index of the multiplet corresponds to the selection of the smallest possible aperture scales

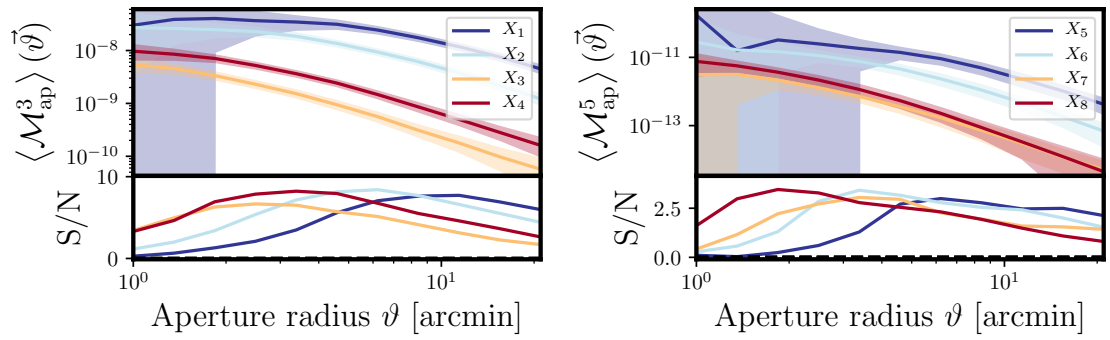


Figure 3.8 Measurements of the unequal radii aperture mass statistics of third (left) and fifth (right) order in the SLICS simulation suite. Each line corresponds to a different set of relative aperture sizes as given in Table 3.1.

| Third order | | Fifth order | |
|---|---|---|---|
| Label | Configuration | Label | Configuration |
| $X_1$ | $(1, 1, 1)$ | $X_5$ | $(1, 1, 1, 1, 1)$ |
| $X_2$ | $(1, 2, 2)$ | $X_6$ | $(1, 1, 2, 2, 2)$ |
| $X_3$ | $(1, 5, 5)$ | $X_7$ | $(1, 1, 5, 5, 5)$ |
| $X_4$ | $(1, 3, 5)$ | $X_8$ | $(1, 2, 3, 4, 5)$ |

Table 3.1 Cofigurations of the aperture radii displayed in Fig. 3.8

from which we then start choosing the next lowest radius in the subsequent dimension up until we reach the combination of the largest possible set of aperture radii - for an example of this path for the third order statistics see Fig. 3.10. Recalling that the second order aperture mass statistic is simply a filtered version of the power spectrum, we should not expect the multiscale extension add any information to that order[12]. For the three statistics we again find a detection significance that is comparable to the equal scale case, meaning that we can extract substantial signal from convergence spectra configurations which are not corresponding to regular polygons. In the right hand side of Fig. 3.9 we plot the joint correlation coefficient of the multiscale cumulants. On the investigated range of scales we only find a slight to modest correlation between the higher order multiscale statistics and the second order one. It also appears that the higher order cumulants exhibit a stronger auto- and cross correlation. However, this is (at least partially) an artefact of the range and sampling density of the chosen radii.

[12] For a different form of the $Q$ filter function like the one proposed in Crittenden et al. (2001) one can easily work this out analytically, see .i.e. Schneider et al. (2005).

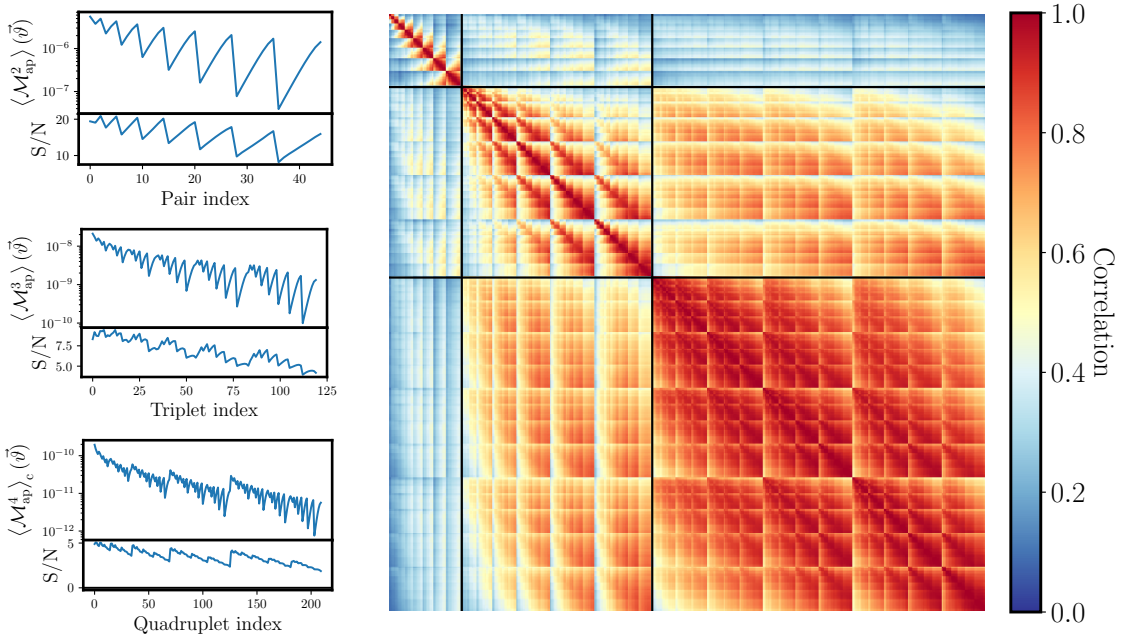


Figure 3.9 Measurements (left) and correlation matrix (right) of the multiscale aperture mass statistics of second, third and fourth order. For each of those statistics we compute all configurations for ten logarithmically spaced radii between $5'$ and $50'$ in which all the apertures have unequal radii. The black lines indicate the blocks of the (cross-) correlations of different orders.

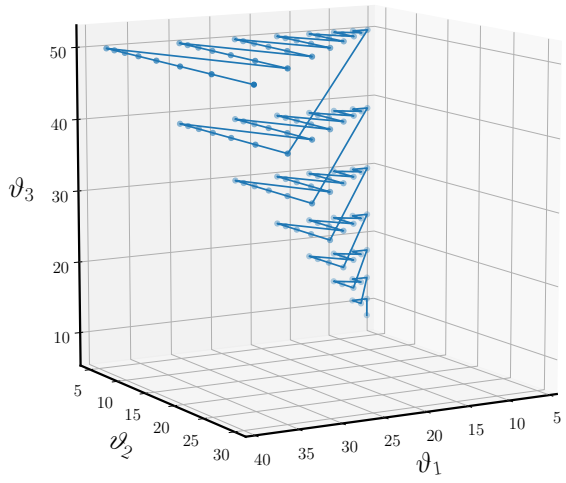


Figure 3.10 Path in which the set of non redundant aperture scales for the third order statistics is traversed. The starting point is the lower corner. The subpath in the $\vartheta_3 = 50'$ plane corresponds to the full path taken for the second order statistics.

## 3.6 Conclusions and Discussion

In this paper we have explored an alternative method for estimating the aperture mass statistics in weak lensing cosmic shear surveys. This study extended our previous work (Porth et al., 2020) in a number of ways: First, we generalized the direct estimator approach to higher statistics, and showed how to rewrite the standard estimator as a product of linear order time sums. Second, we provided the details of the computation of the variance of these estimators. Third, we further generalised the aperture mass statistics to include the multiscale approach. Again, we showed how one could estimate these using linear order products of power sums. The work can be summarised as follows:

In §3.2 we reviewed the background theory of cosmological weak lensing and showed how the connected cumulants of the aperture mass statistics are related to the convergence polyspectra.

In §3.3 we introduced the direct estimator for moments of the aperture mass statistics. We then gave expressions for how the nested sums can be decomposed into a linear combination of products of (multivariate) power sums that facilitates a linearly scaling estimation procedure in the number of galaxies within an aperture. We then generalized this estimator to an ensemble of overlapping apertures and computed its variance. We argued that the aperture cross correlation coefficient leads to a substantial correction to the naive $1/N$ scaling if the apertures are not well separated, and that it also can be used to assess the degree of aperture oversampling that is necessary to capture most of the available information. Finally, we gave a detailed explanation of the algorithms used for our implementation.

In §3.4 we successfully validated our method on Gaussian mock simulations and furthermore verified the linear scaling.

In §3.5 we turned to the SLICS simulation suite and assessed the signal-to-noise of the statistics for a 1000 degree survey following a KiDS-450 like $n(z)$ distribution function. We found that with these specifications significant detections of up to fifth order can be expected for the equal and unequal radii cumulants and that an aperture oversampling rate of at least four extracts nearly all the signal.

In this paper we have neglected the impact of survey masks on the measurement process and the possible bias that this could induce, the exploration of this is sufficient to warrant its own publication and this is the subject of our associated publication (Porth et al. in prep.). Throughout this paper we were mainly concerned with making the extraction of information from higher

order statistics of galaxy shape catalogs computationally feasible and accurate. However, we remained agnostic about further challenges that need to be addressed before applying our methods to real data. For example, one should investigate the required PSF modelling, shape measurement and shear bias calibration quality to not introduce substantial biases in the measurement. Additionally, the range of measurements that can ultimately be used for obtaining cosmological parameter constraints will be limited to the scales for which one can theoretically accurately model those higher order statistics.

## Acknowledgements

LP acknowledges support from a STFC Research Training Grant (grant number ST/R505146/1). RES acknowledges support from the STFC (grant number ST/P000525/1, ST/T000473/1). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National eInfrastructure. We would like to thank Joachim Harnois-Deraps for making public the SLICS mock data, which can be found at http://slics.roe.ac.uk/. This research made use of numpy, a library used for scientific computing and technical computing and matplotlib, a Python library for publication quality graphics (Harris et al., 2020; Hunter, 2007).

## Data Availability

The SLICS mock catalogs are available at https://slics.roe.ac.uk/. Additional data underlying this article will be shared on reasonable request to the corresponding author.

# Appendix

## 3.A Derivations of aperture mass skewness and kurtosis estimators

### 3.A.1 Derivation of the estimator for $\widehat{M_{\rm ap}^3}$

Let us compute the derivation of the skewness $\widehat{M_{\rm ap}^3}$ of the aperture mass. The standard direct estimator is given by:

$$\widehat{M_{\rm ap}^3} = (\pi\vartheta^2)^3 \frac{\sum_{i,j\neq i,k\neq i\neq j}^{N} w_i w_j w_k Q_i Q_j Q_k e_{t,i} e_{t,j} e_{t,k}}{\sum_{i,j\neq i,k\neq i\neq j} w_i w_j w_k} \,. \tag{3.46}$$

It can be shown using the methods described in Schneider et al. (1998) and Porth et al. (2020) that this leads to an unbiased estimator of the skewness. We can rewrite the above estimator by noting that an unconstrained triple sum can be decomposed into the following partial sums:

$$\sum_{i,j,k}^{N} = \sum_{i\neq j\neq k}^{N} + \sum_{i,j=k,i\neq j}^{N} + \sum_{j=k,i,j\neq k}^{N} + \sum_{k=i,j,k\neq i}^{N} + \sum_{i=j=k}^{N} \tag{3.47}$$

This can be rearranged to give:

$$\sum_{i\neq j\neq k}^{N} = \sum_{i,j,k}^{N} - \sum_{i,j=k,i\neq j}^{N} - \sum_{j=k,i,j\neq k}^{N} - \sum_{k=i,j,k\neq i}^{N} - \sum_{i=j=k}^{N} \tag{3.48}$$

Similarly, the unconstrained double sum can be decomposed and rearranged in the following manner:

$$\sum_{i,j} = \sum_{i\neq j} + \sum_{i=j} \;\Rightarrow\; \sum_{i\neq j} = \sum_{i,j} - \sum_{i=j} \,. \tag{3.49}$$

Using this result repeatedly in Eq. (3.48) allows us to rewrite the constrained sums as unconstrained sums:

$$\begin{aligned}\sum_{i\neq j\neq k}^{N} &= \sum_{i,j,k}^{N} - \left(\sum_{i,j=k}^{N} - \sum_{i=j=k}^{N}\right) - \left(\sum_{j,k=i}^{N} - \sum_{i=j=k}^{N}\right) - \left(\sum_{k,i=j}^{N} - \sum_{i=j=k}^{N}\right) - \sum_{i=j=k}^{N} \\ &= \sum_{i,j,k}^{N} - \sum_{i,j=k}^{N} - \sum_{j,k=i}^{N} - \sum_{k,i=j}^{N} + 2\sum_{i=j=k}^{N} = \sum_{i,j,k}^{N}\left[1 - \delta^K_{j,k} - \delta^K_{k,i} - \delta^K_{i,j} + 2\delta^K_{i,j}\delta^K_{i,k}\right] .\end{aligned} \tag{3.50}$$

Hence, on repeatedly using this result we can rewrite the sum in the numerator and denominator of Eq. (3.46) to give us an alternate form for the skewness as:

$$\widehat{M^3_{\rm ap}} = (\pi\vartheta^2)^3 \frac{\left[\sum_{i,j,k}^{N} w_i w_j w_k Q_i Q_j Q_k e_{t,i} e_{t,j} e_{t,k} - 3\sum_{i,j}^{N} w_i w_j^2 Q_i Q_j^2 e_{t,i} e_{t,j}^2 + 2\sum_{i}^{N} w_i^3 Q_i^3 e_{t,i}^3\right]}{\left[\sum_{i,j,k}^{N} w_i w_j w_k - 3\sum_{i,j}^{N} w_i w_j^2 + 2\sum_{i}^{N} w_i^3\right]} . \tag{3.51}$$

If we now divide through each term by $(\sum_i w_i)^3$ and recall expressions Eqs. (3.22) and (3.23) we see that our estimator becomes:

$$\widehat{M^3_{\rm ap}} = \frac{M^3_{\rm s,1} - 3M_{\rm s,2}M_{\rm s,1} + 2M_{\rm s,3}}{1 - 3S_2 + 2S_3} . \tag{3.52}$$

### 3.A.2 Derivation of the estimator for $\widehat{M^4_{\rm ap}}$

The standard direct estimator for the kurtosis of aperture mass is given by:

$$\widehat{M^4_{\rm ap}} = (\pi\vartheta^2)^4 \frac{\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} w_i w_j w_k w_l Q_i Q_j Q_k Q_l e_{t,i} e_{t,j} e_{t,k} e_{t,l}}{\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k} w_i w_j w_k w_l} . \tag{3.53}$$

We follow similar steps to the derivation of the skewness and note that the unconstrained quadruple sum can be written:

$$\begin{aligned}\sum_{i,j,k,l}^{N} = &\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} + \left[\sum_{i,j\neq i,k=l\neq i\neq j}^{N} + 5\text{perms}\right] + \left[\sum_{i=j,k=l\neq i}^{N} + \sum_{i=k,j=l\neq i}^{N} + \sum_{i=l,j=k\neq i}^{N}\right] \\ &+ \left[\sum_{i=j=k,l\neq i}^{N} + \sum_{i=j=l,k\neq i}^{N} + \sum_{i=k=l,j\neq i}^{N} + \sum_{j=k=l,i\neq j}^{N}\right] + \sum_{i=j=k=l}^{N} ,\end{aligned} \tag{3.54}$$

which on rearranging leads us to:

$$\begin{aligned}\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} = &\sum_{i,j,k,l}^{N} - \left[\sum_{i,j\neq i,k=l\neq i\neq j}^{N} + 5\text{perms}\right] - \left[\sum_{i=j,k=l\neq i}^{N} + \sum_{i=k,j=l\neq i}^{N} + \sum_{i=l,j=k\neq i}^{N}\right] \\ &- \left[\sum_{i=j=k,l\neq i}^{N} + \sum_{i=j=l,k\neq i}^{N} + \sum_{i=k=l,j\neq i}^{N} + \sum_{j=k=l,i\neq j}^{N}\right] - \sum_{i=j=k=l}^{N} ,\end{aligned} \tag{3.55}$$

We now make use of our previous results to rewrite the constrained sums on the right-hand side of the expression as unconstrained sums:

$$\begin{aligned}\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} &= \sum_{i,j,k,l}^{N} - \left\{\left[\sum_{i,j,k=l}^{N} - \sum_{i,j=k=l}^{N} - \sum_{j,k=i=l}^{N} - \sum_{k=l,i=j}^{N} + 2\sum_{i=j=k=l}^{N}\right] + 5\text{perms}\right\} \\ &- \left\{\left[\sum_{i=j,k=l}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{i=k,j=l}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{i=l,j=k}^{N} - \sum_{i=j=k=l}^{N}\right]\right\} \\ &- \left\{\left[\sum_{i=j=k,l}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{i=j=l,k}^{N} - \sum_{i=j=k=l}^{N}\right] + \left[\sum_{i=k=l,j}^{N} - \sum_{i=j=k=l}^{N}\right]\right. \\ &\left. + \left[\sum_{k=l=j,i}^{N} - \sum_{i=j=k=l}^{N}\right]\right\} - \sum_{i=j=k=l}^{N} . \end{aligned} \tag{3.56}$$

On making repeated use of the Kroneker delta symbol this can now be compactly written as:

$$\begin{aligned}\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} &= \sum_{i,j,k,l}^{N}\left[1 - \left\{\left[\delta^K_{k,l} - \delta^K_{j,k}\delta^K_{k,l} - \delta^K_{i,k}\delta^K_{k,l} - \delta^K_{i,j}\delta^K_{k,l} + 2\delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + 5\text{perms}\right\}\right. \\ &- \left\{\left[\delta^K_{i,j}\delta^K_{k,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,k}\delta^K_{j,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,l}\delta^K_{j,k} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right\} \\ &- \left\{\left[\delta^K_{i,j}\delta^K_{i,k} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,j}\delta^K_{i,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] + \left[\delta^K_{i,k}\delta^K_{i,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right. \\ &\left.\left. + \left[\delta^K_{j,k}\delta^K_{j,l} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right]\right\} - \delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] . \end{aligned} \tag{3.57}$$

On collecting, cancelling and grouping like terms we see that this can be written:

$$\begin{aligned}\sum_{i,j\neq i,k\neq i\neq j,l\neq i\neq j\neq k}^{N} &= \sum_{i,j,k,l}^{N}\left[1 - \left(\delta^K_{i,j} + \delta^K_{i,k} + \delta^K_{i,l} + \delta^K_{j,k} + \delta^K_{j,l} + \delta^K_{k,l}\right)\right. \\ &+ \left\{\delta^K_{i,j}\delta^K_{k,l} + \delta^K_{i,k}\delta^K_{j,l} + \delta^K_{i,l}\delta^K_{j,k}\right\} + \left\{\delta^K_{i,j}\delta^K_{i,k} + \delta^K_{i,j}\delta^K_{i,l}\right. \\ &\left.\left. + \delta^K_{i,k}\delta^K_{i,l} + \delta^K_{j,i}\delta^K_{j,k} + \delta^K_{j,i}\delta^K_{j,l} + \delta^K_{j,k}\delta^K_{j,l} + \delta^K_{k,i}\delta^K_{k,l} + \delta^K_{k,j}\delta^K_{k,l}\right\} - 6\delta^K_{i,j}\delta^K_{i,k}\delta^K_{i,l}\right] . \end{aligned} \tag{3.58}$$

Hence, on making repeated use of Eq. (3.58) in Eq. (3.53) and along with Eqs. (3.22) and (3.23), the estimator for kurtosis of aperture mass becomes:

$$\widehat{M^4_{\rm ap}} = \frac{M^4_{\rm s,1} - 6M_{\rm s,2}M^2_{\rm s,1} + 3M^2_{\rm s,2} + 8M_{\rm s,3}M_{\rm s,1} - 6M_{\rm s,4}}{1 - 6S_2 + 3S_2^2 + 8S_3 - 6S_4} . \tag{3.59}$$

## 3.B A proof of the general theorem for arbitary order aperture mass statistics

In this section we provide a derivation of the the general form of the $n$-point aperture mass statistic estimator given by Eq. (3.24). At the time of writing, we are not aware that the combinatoric

methods that we have used in the derivation of the general expression have been used before in the cosmological context, and therefore provide a brief overview of them – in particular the Bell polynomials. In what follows we will try to not rely on advanced mathematical methods, but instead use a basic framework to explain how the Bell polynominals are linked to set partitions, and finally how they are connected to the aperture mass estimators.

### 3.B.1 Set partitions and Bell polynomials

We begin by defining a partition $\pi$ of a set $\mathbf{n} = \{1, 2, \cdots, n\}$ as a collection of mutually exclusive subsets (blocks) of $\mathbf{n}$ whose union equals $\mathbf{n}$. In our case all these partitions can be mapped onto an associated partition $\lambda$ being defined as the number of elements of each block in $\pi$. Each element $\lambda$ can be represented as $(n_1, n_2, \cdots, n_m)$ or as $(1^{m_1},\ 2^{m_2},\ \cdots, n^{m_n})$ where for the former expression the $n_i$ denote the length of the $i$th block while for the latter case the $m_i$ represent the number of occurrences of a block of length $i$ in $\pi$. If $\pi$ is a partition of $\mathbf{n}$ having $m$ blocks this implies that $\sum_i m_i = m$ and $\sum_i i\, m_i = n$. We will now show that the following proposition holds:

**Claim**:

For the set $\mathbf{n}$ and a partition $\lambda$ of length $m$ given as $(1^{m_1},\ 2^{m_2},\ \cdots, \ell^{m_\ell})$ there are $\frac{n!}{\prod_{i=1}^{\ell} m_i!(i!)^{m_i}}$ partitions $\pi$ of $\mathbf{n}$ having the same $\lambda(\pi)$.

**Check**:

As a first step we just look at the number of ways the $m$ subsets can be chosen from $\boldsymbol{n}$. This can easily be worked out when noting that for the first subset there are $\binom{n}{n_1}$ choices, for the following $\binom{n-n_1}{n_2}$ etc. Following through all of the subsets we then have

$$\begin{aligned}\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots n_{m-2}}{n_{m-1}}\binom{n_m}{n_m} \\ &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-\cdots n_{m-2})!}{n_{m-1}!n_m!}\frac{n_m!}{n_m!} \\ &= \frac{n!}{n_1!n_2!\cdots n_m!}\end{aligned}$$

possibilities. Shifting this expression to the representation of $\lambda$ given above we see that many of them give the identical partition $\pi$; to get rid of those ones we need to divide by the number of ways all the equal size blocks themselves can be permuted with each other. Applying those conditions we have

$$\frac{1}{\text{Norm.}} \times \frac{n!}{n_1!n_2!\cdots n_m!} = \frac{1}{m_1!m_2!\cdots m_\ell!} \times \frac{n!}{(1!)^{m_1}(2!)^{m_2}\cdots(\ell!)^{m_\ell}} = \frac{n!}{\prod_{i=1}^{\ell} m_i!(i!)^{m_i}} \tag{3.60}$$

possibilities remaining, which is exactly the proposed expression.

With this result in hand we are now in position to understand the form of the partial Bell polynomial $B_{n,m}$ being defined as

$$B_{n,m}(x_1 \cdots, x_{n-m+1}) = \sum_{(m_1,\cdots,m_{n-m+1})\in P_{n,m}} \frac{n!}{m_1!\cdots m_{n-m+1}!}\left(\frac{x_1}{1!}\right)^{m_1}\cdots\left(\frac{x_{n-m+1}}{(n-m+1)!}\right)^{m_{n-m+1}} , \tag{3.61}$$

where

$$P_{n,m} \equiv \left\{(m_1,\cdots m_{n-m+1}) \in \mathbb{N}_0^{n-m+1} \mid \sum_{i=1}^{n-m+1} m_i = m \ , \ \sum_{i=1}^{n-m+1} i\, m_i = n\right\} .$$

Comparing the prefactors and the index set[13] with our discussion above we see that the partial Bell polynomials simply sum over all the partitions $\lambda$ of $\mathbf{n}$ having a fixed $m$, i.e. they list the number of ways a set consisting the $n$ objects can be partitioned into $m$ blocks. For example, looking at $B_{4,2}$ the allowed index combinations are $\{(0,2,0),(1,0,1)\}$ such that Eq. (3.61) evaluates to $B_{4,2} = 4x_1x_3 + 3x_2^2$. We note in the passing that these expressions generate the same prefactors that arise in the halo model, i.e. we can relate the structure of $B_{4,2}$ to the two-halo term of the halo model trispectrum.

Finally, we define the complete Bell polynominal $B_n$ which list all possible partitions of $n$ objects:

$$B_n(x_1\cdots,x_{n-m+1}) = \sum_{m=1}^{n} B_{n,m}(x_1\cdots,x_{n-m+1}) = \sum_{m=1}^{n}\sum_{\pi\in P_{n,m}}\prod_{i=1}^{n-m(\lambda(\pi))+1} x_i^{m_i(\lambda(\pi))} , \tag{3.62}$$

where the first equality states the formal definition and the second one rewrites it into an explicit sum over all the partitions of the set $\boldsymbol{n}$.

### 3.B.2 Sums over unequal indices and Bell polynomials

Let us look at the simple expression $\sum_{i=1}^{N}\sum_{j\neq i}^{N} x_i x_j$. A naive implementation of this double sum would imply a quadratic complexity of the corresponding program. A much faster way resulting in linear complexity can be achieved when noting that $\left(\sum_{i_1=1}^{N} x_{i_1}\right)\left(\sum_{i_2=1}^{N} x_{i_2}\right) = \sum_{i_1=1}^{N}\sum_{i_2\neq i_1}^{N} x_{i_1}x_{i_2} + \sum_{i_1=1}^{N} x_1^2$ . We can easily generalize this pattern by treating the number of indices as the set $\mathbf{n}$ from the previous subsection. Then all the different partitions $\lambda$ of this set correspond to different ways these indices can be set equal with one another; the corresponding

[13] The upper limit is given by the partition having the largest possible block size, namely $(1^{m-1},\ 2^0,\cdots,\ (n-(m-1))^1)$

prefactors can be obtained via the Bell polynominal. To clarify this statement we write down as an example the expression for $n = 4$:

$$\begin{aligned}\left(\sum_i x_i\right)^4 &= \sum_{i_1\neq i_2\neq i_3\neq i_4} x_{i_1}x_{i_2}x_{i_3}x_{i_4} + \left(\sum_{i_1\neq i_3\neq i_4} x_{i_1}^2 x_{i_3}x_{i_4} + 5\text{perm.}\right) \\ &\quad + \left(\sum_{i_1\neq i_2} x_{i_1}^3 x_{i_2} + 3\text{perm.}\right) + \left(\sum_{i_1\neq i_3} x_{i_1}^2 x_{i_3}^2 + 2\text{perm.}\right) + \sum_{i_1} x_{i_1}^4 \\ &\sim \quad (1^4,\, 2^0\,,\, 3^0\,,\, 4^0) \;+\; 6\times(1^2\,,\, 2^1\,,\, 3^0\,,\, 4^0) \\ &\quad + \;4\times(1^1\,,\, 2^0\,,\, 3^1\,,\, 4^0) \;+\; 3\times(1^0\,,\, 2^2\,,\, 3^0\,,\, 4^0) \;+\; (1^0\,,\, 2^0\,,\, 3^0\,,\, 4^1)\end{aligned} \tag{3.63}$$

From here we see that we can express a sum over $n$ unequal indices in terms of two power sums and a set of related sums over at most $n-1$ unequal indices. Repeating the same argument on the latter sums one eventually arrives at an expression only involving power sums. Carrying out aforementioned calculations for our example this yields

$$\begin{aligned}\sum_{i_1\neq i_2\neq i_3\neq i_4} x_{i_1}x_{i_2}x_{i_3}x_{i_4} &= \left(\sum_i x_i\right)^4 - 6\left(\sum_i x_i\right)^2\left(\sum_i x_i^2\right) + 8\left(\sum_i x_i\right)\left(\sum_i x_i^3\right) \\ &\quad + 3\left(\sum_i x_i^2\right)^2 - 6\left(\sum_i x_i^4\right)\,.\end{aligned} \tag{3.64}$$

Comparing the latter two expressions we note that their index partitions are the same, but that they differ in some signs and prefactors; namely there is a negative sign for an odd partition length $m$ and an additional multiplicative factor of $(i-1)!$ for each block of length $i$. Looking at the structure of Eq. (3.64), i.e. the fact that all of its summands correspond to a partition of an integer set and that furthermore it constitutes of $n$ different building blocks we might be tempted to cast it in terms of Bell polynomials with the identifying the $x_\ell$ from Eq. (3.62) with the power sums $c_\ell \sum_i x_i^\ell \;:\; c_\ell \in \mathbb{R}$. In the next paragraphs we formalize these observations and from there determine the $c_\ell$.

The first difference can be motivated most easily by choosing a graphical representation in which we draw each index as a single point. Then the prefactors in Eq. (3.63) are given by the number of ways one can group together different points such that they constitute the corresponding partition whereas for Eq. (3.64) it additionally matters in which order these points have been set equal with each other, which in mathematical terms is described by how many closed cycles one can draw between them. The induced correction of $(\ell-1)!$ for a block of length $\ell$ can be absorbed in the Bell polynomial by setting $c_\ell = (\ell-1)!$.

The second observation can be generalized inductively. Looking at our example of $n = 4$ we see that the sign for each partition $\lambda$ is given by $\text{sgn}(\lambda) = \prod_{i=1}^n (-1)^{m_i(\lambda)((i+1)\text{mod}2)}$, that is each

block of even length contributes a negative sign. Performing the induction step we have

$$\sum_{i_1\neq\cdots i_n\neq i_{n+1}} x_{i_1}\cdots x_{i_{n+1}} = \left(\sum_{i_{n+1}} x_{i_{n+1}}\right)\left(\sum_{i_1\neq\cdots i_3\neq i_n} x_{i_1}\cdots x_{i_n}\right) - \left[\left(\sum_{i_1\neq\cdots i_3\neq i_n} x_{i_1}^2 x_{i_2}\cdots x_{i_n}\right) + (n-1)\text{ perm.}\right] . \tag{3.65}$$

Looking at the modification of the partitions, for the first term we have $m_1 \to m_1 + 1$ for all $\lambda$ such that we would not have expected any sign flips. For the second term, we need to update the block in which the identical index sits, assuming it had length $k$ we have $m_k \to m_k - 1$ and $m_{k+1} \to m_{k+1}+1$. In case of an even $k$ reducing its occurence by one induces an additional sign flip whereas for odd $k$s we get a sign flip for the increase of $m_{k+1}$. Putting things together we conclude that we could predict the correct signs by examining the partition structures. Therefore, setting $c_\ell = (-1)^{(\ell+1) \bmod 2}(\ell-1)!$ in Eq. (3.62) will reproduce generalizations of Eq. (3.64). We can brush this in a nicer shape by setting $c_\ell = -(\ell-1)!$ and furthermore multiplying $B_n$ by $(-1)^n$; this modification effectively just multiplies each term of the previous result by an even power of negative one.

With these two modifications in hand we can finally write down the main result of this subsection, namely the way on how to transform a sum over unequal indices into a sum over products of power sums:

$$\sum_{i_1\neq\cdots\neq i_n} x_{i_1}\cdots x_{i_n} = (-1)^n\, B_n\left(-0!\sum_i x_i,\ -1!\sum_i x_i^2, \cdots,\ -(n-1)!\sum_i x_i^n\right) \tag{3.66}$$

### 3.B.3 Application to the aperture mass estimator

Looking at the form of Eq. (3.66), the expression for the direct estimator of the aperture statistics with equal aperture radii Eq. (3.24) immediately follows when identifying the arguments in the nominator and denominator with the power sums $M_{\mathrm{s},m}$ and $S_m$ and cancelling the overall sign.

For the case of unequal aperture radii we still need to do a bit more work. Looking back to our previous example Eq. (3.63), having unequal aperture radii induces different values of the $Q$ filters such that the $x_i$ cannot be taken to be the same variable anymore. Hence we have to replace the prefactors in Eq. (3.63) by a sum over all the possible ways the different radii can be partitioned. The second set of prefactors that arises when going to Eq. (3.64) still applies in the case of unequal radii as it effectively corresponds to swapping two aperture radii in the

corresponding multivariate power sum Eq. (3.31). Thus it seems appropriate to formulate the solution via summing over partitions, such that we can rewrite Eq. (3.29) as

$$\widehat{M^n_{\rm ap}}(R_1,...,R_n) = \frac{\sum_{m=1}^{n}\sum_{\pi\in P_{n,m}}(-1)^m\prod_{i=1}^{m}(n_i-1)!\,\hat{M}^{(n_i)}_{s,(\alpha_1(\pi_i),\cdots,\alpha_n(\pi_i))}}{\sum_{m=1}^{n}\sum_{\pi\in P_{n,m}}(-1)^m\prod_{i=1}^{m}(n_i-1)!\,\hat{S}^{(n_i)}_{(\alpha_1(\pi_i),\cdots,\alpha_n(\pi_i))}}\,. \tag{3.67}$$

We note that from this formulation one can build an efficient way of computing Eq. (3.32) within the subset $\mathcal{U}$ of the datacube $[R_1,\cdots R_m]^n \;\; (m\geq n)$ in which neither of the indices are equal: This is due to the fact that the number of power sums in which $1\leq i\leq n$ radii are selected is simply given by $\binom{m}{i}$ and therefore the n-dimensional hypercube of aperture radii can be constructed from a set consisting of just $\sum_{i=1}^{n}\binom{m}{i}$ power sums. After allocating those power sums for all the galaxies within an aperture we can then enumerate through the relevant aperture radii multiplets, select the relevant subsets of the power sums, and then again apply the transformation equation (3.32) to transform to the multiscale aperture mass moments, or equivalently to their corresponding connected parts. With the help of this procedure we were able to conduct the full analysis displayed in Fig. 3.9 on the SLICS ensemble (a total of around 2.5 billion galaxies) within just 6000 CPU hours.

### 3.B.4 Expressions of the accelerated estimator for low orders (unequal radii)

In order to save space we only write down the expressions for the nominator or Eq. (3.32), the denominator will have an identical structure. As expected, the number of sums in the $n$th order estimator equals the $n$th Bell number.

$$
\begin{aligned}
\widehat{M^1_{\mathrm{ap}}}(\vartheta_1) &= \frac{1}{\mathrm{norm}} \times M^{(1)}_{\mathrm{s},1} \; ; \\
\widehat{M^2_{\mathrm{ap}}}(\vartheta_1,\vartheta_2) &= \frac{1}{\mathrm{norm}} \times \left\{ M^{(1)}_{\mathrm{s},1,0} M^{(1)}_{\mathrm{s},0,1} - M^{(2)}_{\mathrm{s},1,1} \right\} \; ; \\
\widehat{M^3_{\mathrm{ap}}}(\vartheta_1,\vartheta_2,\vartheta_3) &= \frac{1}{\mathrm{norm}} \times \left\{ M^{(1)}_{\mathrm{s},1,0,0} M^{(1)}_{\mathrm{s},0,1,0} M^{(1)}_{\mathrm{s},0,0,1} - \left[ M^{(2)}_{\mathrm{s},1,1,0}\, M^{(1)}_{\mathrm{s},0,0,1} \;+\; 2 \text{ perm.} \right] + 2\, M^{(3)}_{\mathrm{s},1,1,1} \right\} \; ; \\
\widehat{M^4_{\mathrm{ap}}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4) &= \frac{1}{\mathrm{norm}} \times \Big\{ M^{(1)}_{\mathrm{s},1,0,0,0} M^{(1)}_{\mathrm{s},0,1,0,0} M^{(1)}_{\mathrm{s},0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,1} - \left[ M^{(2)}_{\mathrm{s},1,1,0,0} M^{(1)}_{\mathrm{s},0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,1} \;+\; 5 \text{ perm.} \right] \\
&\qquad + \left[ M^{(2)}_{\mathrm{s},1,1,0,0} M^{(2)}_{\mathrm{s},0,0,1,1} \;+\; 2 \text{ perm.} \right] + 2 \left[ M^{(3)}_{\mathrm{s},1,1,1,0} M^{(1)}_{\mathrm{s},0,0,0,1} \;+\; 3 \text{ perm.} \right] - 6\, M^{(4)}_{\mathrm{s},1,1,1,1} \Big\} \; ; \\
\widehat{M^5_{\mathrm{ap}}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4,\vartheta_5) &= \frac{1}{\mathrm{norm}} \times \Big\{ M^{(1)}_{\mathrm{s},1,0,0,0,0} M^{(1)}_{\mathrm{s},0,1,0,0,0} M^{(1)}_{\mathrm{s},0,0,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1} \\
&\qquad - \left[ M^{(2)}_{\mathrm{s},1,1,0,0,0} M^{(1)}_{\mathrm{s},0,0,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1} \;+\; 9 \text{ perm.} \right] \\
&\qquad + \left[ M^{(2)}_{\mathrm{s},1,1,0,0,0} M^{(2)}_{\mathrm{s},0,0,1,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1} \;+\; 14 \text{ perm.} \right] \\
&\qquad + 2 \left[ M^{(3)}_{\mathrm{s},1,1,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1} + 9 \text{ perm.} \right] - 2 \left[ M^{(3)}_{\mathrm{s},1,1,1,0,0} M^{(2)}_{\mathrm{s},0,0,0,1,1} \;+\; 9 \text{ perm.} \right] \\
&\qquad + 6 \left[ M^{(4)}_{\mathrm{s},1,1,1,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1} \;+\; 4 \text{ perm.} \right] + 24\, M^{(5)}_{\mathrm{s},1,1,1,1,1} \Big\} \; ; \\
\widehat{M^6_{\mathrm{ap}}}(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4,\vartheta_5,\vartheta_6) &= \frac{1}{\mathrm{norm}} \times \Big\{ M^{(1)}_{\mathrm{s},1,0,0,0,0,0} M^{(1)}_{\mathrm{s},0,1,0,0,0,0} M^{(1)}_{\mathrm{s},0,0,1,0,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,0,1}
\end{aligned}
$$

$$
\begin{aligned}
&- \left[ M^{(2)}_{\mathrm{s},1,1,0,0,0,0} M^{(1)}_{\mathrm{s},0,0,1,0,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} \;+\; 14 \text{ perm.} \right] \\
&+ \left[ M^{(2)}_{\mathrm{s},1,1,0,0,0,0} M^{(2)}_{\mathrm{s},0,0,1,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,0,1} \;+\; 44 \text{ perm.} \right] \\
&- \left[ M^{(2)}_{\mathrm{s},1,1,0,0,0,0} M^{(2)}_{\mathrm{s},0,0,1,1,0,0} M^{(2)}_{\mathrm{s},0,0,0,0,1,1} \;+\; 14 \text{ perm.} \right] \\
&+ 2 \left[ M^{(3)}_{\mathrm{s},1,1,1,0,0,0} M^{(1)}_{\mathrm{s},0,0,0,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} + 19 \text{ perm.} \right] \\
&- 2 \left[ M^{(3)}_{\mathrm{s},1,1,1,0,0,0} M^{(2)}_{\mathrm{s},0,0,0,1,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} \;+\; 59 \text{ perm.} \right] \\
&+ 4 \left[ M^{(3)}_{\mathrm{s},1,1,1,0,0,0} M^{(3)}_{\mathrm{s},0,0,0,1,1,1} \;+\; 9 \text{ perm.} \right] \\
&- 6 \left[ M^{(4)}_{\mathrm{s},1,1,1,1,0,0} M^{(1)}_{\mathrm{s},0,0,0,0,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,0,1} \;+\; 14 \text{ perm.} \right] \\
&+ 6 \left[ M^{(4)}_{\mathrm{s},1,1,1,1,0,0} M^{(2)}_{\mathrm{s},0,0,0,0,1,1} \;+\; 14 \text{ perm.} \right] + 24 \left[ M^{(5)}_{\mathrm{s},1,1,1,1,1,0} M^{(1)}_{\mathrm{s},0,0,0,0,0,1} \;+\; 5 \text{ perm.} \right] \\
&- 120 \, M^{(6)}_{\mathrm{s},1,1,1,1,1,1} \Big\}
\end{aligned}
$$

## 3.C Variance of the direct estimator

### 3.C.1 Motivation of the shape and multiplicity factor

We recall the definition of the $\widehat{M_{\rm ap}^n}$ variance:

$$\begin{aligned}\sigma^2\left[\widehat{M_{\rm ap}^n}\right] &= \mathbb{E}\left[\left(\widehat{M_{\rm ap}^n}\right)^2\right] - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 \\ &= \frac{\left(\pi\vartheta^2\right)^{2n}}{\left(\sum_{\neq} w_{j_1}\cdots w_{j_n}\right)^2}\cdot\mathbb{E}\left[\sum_{\neq} w_{i_1}\cdots w_{i_n}x_{i_1}\cdots x_{i_n}\cdot\sum_{\neq} w_{j_1}\cdots w_{j_n}x_{j_1}\cdots x_{j_n}\right] \\ &\quad - \left\langle \mathcal{M}_{\rm ap}^n\right\rangle^2 , \end{aligned} \tag{3.68}$$

where we defined $x_i \equiv Q_i e_{t,i}$ for notational simplicity. We proceed along the standard lines by decomposing the expectation value in an averaging step $A$ over the intrinsic ellipticity distribution, another one $P$ over the galaxy positions, and finally one over the cosmological ensemble. Let us start by applying the ellipticity averaging procedure for which $A(e_i,e_j) \equiv \frac{\sigma_\epsilon^2}{2}\delta_{i,j}^K + \gamma_i\gamma_j\left(1-\delta_{i,j}^K\right)$. Noting that each summation sign in (3.68) runs over an index set where all the indices are unequal, we see that only indices between the two sums can be contracted to yield the shape noise expression. We can represent the index structure graphically as $|\, i_1\ \cdots\ i_n \,|\, j_1\ \cdots\ j_n|$ and define a contraction as a line between two indices of the $i$ and $j$ set. The prefactor of the term in the $A$-averaging is then given by the number of possible contractions.

As an example, let us compute the prefactor when applying two contractions in the variance of the third order statistics. For the first contraction there are 9 possibilities, while for each second one there are only for indices remaining, giving 4 further possibilities. As the contractions are interchangeable we need to divide the result by two to yield a prefactor of 18. A graphical representation of this explanation would look as follows:

$$\left|\overbrace{i_1\ \overbrace{i_2\ i_3}\ |\ j_1\ \overbrace{j_2\ j_3}}\right| = \frac{9}{2}\times\left|\overbrace{i_2\ i_3}\ |\ \overbrace{j_2\ j_3}\right| = \frac{9\cdot 4}{2!} = 18\,.$$

This scheme allows us to easily generalize our example to performing $\ell$ contractions on the $n$th order statistics, giving a prefactor of $C_2(n,\ell) \equiv \frac{n^2(n-1)^2\cdots(n-\ell-1)^2}{\ell!}$ .

For the position averaging we can repeat the same argument, as $P(Q_i\gamma_i Q_j\gamma_j) \sim M_{\rm s,2}\delta_{i,j}^K + \mathcal{M}_{\rm ap}^2\left(1-\delta_{i,j}^K\right)$. If we already have performed $\ell$ contractions for the $A$-averaging, there are only $(n-\ell)$ free indices left in each block - hence there will be $C_2(n-\ell,p)$ possibilities to perform $p$ additional contractions in the $P$-averaging.

Next we compute the expectation value for a given index set in which we have performed $\ell$

contractions in the $A$-averaging and $p$ contractions in the $P$-averaging:

$$
\begin{aligned}
&\Bigg\langle P \Bigg( \sum_{\neq} w_{i_1}^2 Q_{i_1}^2 \cdots w_{i_\ell}^2 Q_{i_\ell}^2 \; w_{i_{\ell+1}}^2 Q_{i_{\ell+1}}^2 \gamma_{t,i_{\ell+1}}^2 \cdots w_{i_{\ell+p}}^2 Q_{i_{\ell+p}}^2 \gamma_{t,i_{\ell+p}}^2 \\
&\qquad w_{i_{\ell+p+1}} Q_{i_{\ell+p+1}} \gamma_{t,i_{\ell+p+1}} \cdots w_{i_n} Q_{i_n} \gamma_{t,i_n} \; w_{j_{\ell+p+1}} Q_{j_{\ell+p+1}} \gamma_{t,j_{\ell+p+1}} \cdots w_{j_n} Q_{i_n} \gamma_{t,j_n} \Bigg) \Bigg\rangle \\
&\equiv \Bigg\langle \prod_{i=1}^{N} \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} \sum_{\neq} w_{i_1}^2 Q_{i_1}^2 \cdots w_{i_\ell}^2 Q_{i_\ell}^2 \; w_{i_{\ell+1}}^2 Q_{i_{\ell+1}}^2 \gamma_{t,i_{\ell+1}}^2 \cdots w_{i_{\ell+p}}^2 Q_{i_{\ell+p}}^2 \gamma_{t,i_{\ell+p}}^2 \\
&\qquad w_{i_{\ell+p+1}} Q_{i_{\ell+p+1}} \gamma_{t,i_{\ell+p+1}} \cdots w_{i_n} Q_{i_n} \gamma_{t,i_n} \; w_{j_{\ell+p+1}} Q_{j_{\ell+p+1}} \gamma_{t,j_{\ell+p+1}} \cdots w_{j_n} Q_{i_n} \gamma_{t,j_n} \Bigg\rangle \\
&= \sum_{\neq} w_{i_1}^2 \cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}} \cdots w_{i_n} w_{j_{\ell+p+1}} \cdots w_{j_n} \left( \prod_{i \in \{i_1, \cdots, i_\ell\}} \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} Q_i^2 \right) \\
&\qquad \times \left\langle \left( \prod_{i \in \{i_{\ell+1}, \cdots, i_{\ell+p}\}} \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} Q_i^2 \gamma_{t,i}^2 \right) \left( \prod_{i \in \{i_{\ell+p+1}, \cdots, j_n\}} \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} Q_i \gamma_{t,i} \right) \right\rangle \left( \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} \right)^{N-2(\ell+p)} \\
&= \sum_{\neq} w_{i_1}^2 \cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}} \cdots w_{i_n} w_{j_{\ell+p+1}} \cdots w_{j_n} \; \times \prod_{i=1}^{\ell} \left( \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_i}{\pi\vartheta^2} Q_i^2 \right) \\
&\qquad \times \left\langle \prod_{j=1}^{p} \left( \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_j}{\pi\vartheta^2} Q_j^2 \gamma_j^2 \right)^{2(n-\ell-p)} \prod_{k=1}^{2(n-\ell-p)} \left( \int_{\mathrm{Ap.}} \frac{\mathrm{d}^2\theta_k}{\pi\vartheta^2} Q_k \gamma_k \right) \right\rangle \\
&\equiv \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}} \cdots w_{i_n} w_{j_{\ell+p+1}} \cdots w_{j_n}}{(\pi\vartheta)^{2n}} \; \times \; M_{g,2}^{\ell} \langle \mathcal{M}_{\mathrm{s},2}^{p} \mathcal{M}_{\mathrm{ap}}^{2(n-\ell-p)} \rangle \, .
\end{aligned}
$$

Note that in this derivation the order of the contracted indices does not matter as they all end up to be integration variables. If we now combine this result together with the multiplicity factors we can write a closed form expression for (3.68):

$$
\begin{aligned}
\sigma^2 \left[ \widehat{M_{\mathrm{ap}}^n} \right] &= \sum_{\ell=0}^{n} C_2(n,\ell) \left( \frac{\sigma_\epsilon^2}{2} \right)^{\ell} M_{g,2}^{\ell} \sum_{p=0}^{n-\ell} \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_{\ell+p}}^2 w_{i_{\ell+p+1}} \cdots w_{i_n} w_{j_{\ell+p+1}} \cdots w_{j_n}}{\left( \sum_{\neq} w_{i_1} \cdots w_{i_n} \right)^2} \\
&\qquad \times C_2(n-\ell,p) \langle \mathcal{M}_{\mathrm{s},2}^{p} \mathcal{M}_{\mathrm{ap}}^{2(n-\ell-p)} \rangle \; - \; \left\langle \mathcal{M}_{\mathrm{ap}}^{n} \right\rangle^2 \\
&\approx \sum_{\ell=0}^{n} \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_\ell}^2 w_{i_{\ell+1}} \cdots w_{i_n} w_{j_{\ell+1}} \cdots w_{j_n}}{\left( \sum_{\neq} w_{i_1} \cdots w_{i_n} \right)^2} \\
&\qquad \times \ell! \binom{n}{\ell} \binom{n}{\ell} \left( \frac{\sigma_\epsilon^2}{2} \right)^{\ell} M_{g,2}^{\ell} \left\langle \mathcal{M}_{\mathrm{ap}}^{2(n-\ell)} \right\rangle \; - \; \left\langle \mathcal{M}_{\mathrm{ap}}^{n} \right\rangle^2 \\
&\approx n! \frac{\sum_{\neq} w_{i_1}^2 \cdots w_{i_n}^2}{\left( \sum_{\neq} w_{i_1} \cdots w_{i_n} \right)^2} \left( \frac{\sigma_\epsilon^2}{2} \right)^{n} M_{g,2}^{n} \, .
\end{aligned}
\tag{3.69}
$$

The first line is equivalent to (3.35) when combining the multiplicity factors and adjusting the indices. The second line makes the approximation that each of the $\mathcal{M}_{\mathrm{s},2}$ are negligible (which is true for large $N$); for the final line we only keep the shot noise contribution.

### 3.C.2 Modifications for unequal aperture radii

In case of multiple apertures the structure of the variance is basically unchanged, the only thing we need to adjust is to use the multivariate version of the power sums and to replace the multiplicity factor with a sum over the actual multivariate expressions such that their radii correspond to the structure of the contracted indices. If we then take the shot noise dominated case we end up with:

$$\sigma^2_{\text{shot}}\left[\widehat{M^n_{\text{ap}}}(R_1,\cdots,R_n)\right] = \frac{\sum'_{\neq} w^2_{i_1}\cdots w^2_{i_n}}{\left(\sum'_{\neq} w_{i_1}\cdots w_{i_n}\right)^2}\left(\frac{\sigma^2_\epsilon}{2}\right)^n \times \sum_{\alpha_1\neq\cdots\neq\alpha_n}\prod_{i=1}^{n} G_2\left(\frac{\max\left(\{R_i,R_{\alpha_i}\}\right)}{\min\left(\{R_i,R_{\alpha_i}\}\right)}\right), \tag{3.70}$$

where we define $G_2$ as the multiple radii generalization of $\mathcal{M}_{\text{g},2}$:

$$G_2(\alpha) \equiv \pi R^2 \int \mathrm{d}^2\boldsymbol{\theta}\, Q_R(\theta) Q_{\alpha R}(\theta) = \frac{72}{\alpha^2}\left[\frac{1}{24} - \frac{1}{8\alpha^2} + \frac{1}{10\alpha^4}\right] \quad (\alpha \geq 1)$$

where the second equality denotes the corresponding equation for the polynomial filter. Note that for the corresponding inverse shot noise weighting scheme only the sum over the weights matters, as the remainder of the above expression is constant and can be factored out.

### 3.C.3 Explicit expressions for low orders

Here we collect the lowest order explicit expressions for (3.35). The second order expression was first derived in (Schneider, 1998). Note that our prefactors differ from the ones defined in (Munshi & Coles, 2003).

$$\sigma^2\left[\widehat{M^1_{\rm ap}}\right] = \frac{1}{\left(\sum_{\neq} w_{i_1}\right)^2}\left\{\sum_{\neq} w_{i_1}w_{j_1}\left\langle\mathcal{M}^2_{\rm ap}\right\rangle + \sum_{\neq} w^2_{i_1}\left\langle\mathcal{M}_{s,2}\right\rangle + 1\mathcal{M}_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)\sum_{\neq} w^2_{i_1}\right\} - \left\langle\mathcal{M}_{\rm ap}\right\rangle^2$$

$$\begin{aligned}\sigma^2\left[\widehat{M^2_{\rm ap}}\right] = \frac{1}{\left(\sum_{\neq} w_{i_1}w_{i_2}\right)^2}\Bigg\{&\sum_{\neq} w_{i_1}w_{j_1}w_{i_2}w_{j_2}\left\langle\mathcal{M}^4_{\rm ap}\right\rangle + 4\sum_{\neq} w^2_{i_1}w_{i_2}w_{j_2}\left\langle\mathcal{M}_{s,2}\mathcal{M}^2_{\rm ap}\right\rangle + 2\sum_{\neq} w^2_{i_1}w^2_{i_2}\left\langle\mathcal{M}^2_{s,2}\right\rangle\\ &+4\mathcal{M}_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)\left[\sum_{\neq} w^2_{i_1}w_{i_2}w_{j_2}\left\langle\mathcal{M}^2_{\rm ap}\right\rangle + \sum_{\neq} w^2_{i_1}w^2_{i_2}\left\langle\mathcal{M}_{s,2}\right\rangle\right] + 2\mathcal{M}^2_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)^2\sum_{\neq} w^2_{i_1}w^2_{i_2}\Bigg\} - \left\langle\mathcal{M}^2_{\rm ap}\right\rangle^2\end{aligned}$$

$$\begin{aligned}\sigma^2\left[\widehat{M^3_{\rm ap}}\right] = \frac{1}{\left(\sum_{\neq} w_{i_1}w_{i_2}w_{i_3}\right)^2}\Bigg\{&\sum_{\neq} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}^6_{\rm ap}\right\rangle + 9\sum_{\neq} w^2_{i_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}_{s,2}\mathcal{M}^4_{\rm ap}\right\rangle\\ &+18\sum_{\neq} w^2_{i_1}w^2_{i_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}^2_{s,2}\mathcal{M}^2_{\rm ap}\right\rangle + 6\sum_{\neq} w^2_{i_1}w^2_{i_2}w^2_{i_3}\left\langle\mathcal{M}^3_{s,2}\right\rangle + 9\mathcal{M}_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)\left[\sum_{\neq} w^2_{i_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}^4_{\rm ap}\right\rangle\right.\\ &\left.+4\sum_{\neq} w^2_{i_1}w^2_{i_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}_{s,2}\mathcal{M}^2_{\rm ap}\right\rangle + 2\sum_{\neq} w^2_{i_1}w^2_{i_2}w^2_{i_3}\left\langle\mathcal{M}^2_{s,2}\right\rangle\right] + 18\mathcal{M}^2_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)^2\left[\sum_{\neq} w^2_{i_1}w^2_{i_2}w_{i_3}w_{j_3}\left\langle\mathcal{M}^2_{\rm ap}\right\rangle\right.\\ &\left.+\sum_{\neq} w^2_{i_1}w^2_{i_2}w^2_{i_3}\left\langle\mathcal{M}_{s,2}\right\rangle\right] +6\mathcal{M}^3_{g,2}\left(\frac{\sigma^2_\epsilon}{2}\right)^3\sum_{\neq} w^2_{i_1}w^2_{i_2}w^2_{i_3}\Bigg\} - \left\langle\mathcal{M}^3_{\rm ap}\right\rangle^2\end{aligned}$$

$$
\begin{aligned}
\sigma^2\left[\widehat{M_{\text{ap}}^4}\right] = & \frac{1}{\left(\sum_{\neq} w_{i_1}w_{i_2}w_{i_3}w_{i_4}\right)^2}\left\{\sum_{\neq} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{\text{ap}}^8\right\rangle + 16\sum_{\neq} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}\mathcal{M}_{\text{ap}}^6\right\rangle\right. \\
& + 72\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}^2\mathcal{M}_{\text{ap}}^4\right\rangle + 96\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}^3\mathcal{M}_{\text{ap}}^2\right\rangle + 24\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle \mathcal{M}_{s,2}^4\right\rangle \\
& + 16\mathcal{M}_{g,2}\left(\frac{\sigma_\epsilon^2}{2}\right)\left[\sum_{\neq} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{\text{ap}}^6\right\rangle + 9\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}\mathcal{M}_{\text{ap}}^4\right\rangle\right. \\
& \left. + 18\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}^2\mathcal{M}_{\text{ap}}^2\right\rangle + 6\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle \mathcal{M}_{s,2}^3\right\rangle\right] + 72\mathcal{M}_{g,2}^2\left(\frac{\sigma_\epsilon^2}{2}\right)^2 \\
& \left[\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}\left\langle \mathcal{M}_{\text{ap}}^4\right\rangle + 4\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle \mathcal{M}_{s,2}\mathcal{M}_{\text{ap}}^2\right\rangle + 2\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle \mathcal{M}_{s,2}^2\right\rangle\right] \\
& \left. + 96\mathcal{M}_{g,2}^3\left(\frac{\sigma_\epsilon^2}{2}\right)^3\left[\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}\left\langle \mathcal{M}_{\text{ap}}^2\right\rangle + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\left\langle \mathcal{M}_{s,2}\right\rangle\right] + 24\mathcal{M}_{g,2}^4\left(\frac{\sigma_\epsilon^2}{2}\right)^4\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2\right\} - \left\langle \mathcal{M}_{\text{ap}}^4\right\rangle^2 \\
\sigma^2\left[\widehat{M_{\text{ap}}^5}\right] = & \frac{1}{\left(\sum_{\neq} w_{i_1}w_{i_2}w_{i_3}w_{i_4}w_{i_5}\right)^2}\left\{\sum_{\neq} w_{i_1}w_{j_1}w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle \mathcal{M}_{\text{ap}}^{10}\right\rangle\right. \\
& + 25\sum_{\neq} w_{i_1}^2 w_{i_2}w_{j_2}w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle \mathcal{M}_{s,2}\mathcal{M}_{\text{ap}}^8\right\rangle + 200\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}w_{j_3}w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle \mathcal{M}_{s,2}^2\mathcal{M}_{\text{ap}}^6\right\rangle \\
& \left. + 600\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}w_{j_4}w_{i_5}w_{j_5}\left\langle \mathcal{M}_{s,2}^3\mathcal{M}_{\text{ap}}^4\right\rangle + 600\sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}w_{j_5}\left\langle \mathcal{M}_{s,2}^4\mathcal{M}_{\text{ap}}^2\right\rangle\right.
\end{aligned}
$$

$$
\begin{aligned}
&+ 120 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \left\langle \mathcal{M}_{s,2}^5 \right\rangle + 25 \mathcal{M}_{g,2} \left( \frac{\sigma_\epsilon^2}{2} \right) \left[ \sum_{\neq} w_{i_1}^2 w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{\text{ap}}^8 \right\rangle \right. \\
&+ 16 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^6 \right\rangle + 72 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^4 \right\rangle \\
&\left. + 96 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\text{ap}}^2 \right\rangle + 24 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \left\langle \mathcal{M}_{s,2}^4 \right\rangle \right] + 200 \mathcal{M}_{g,2}^2 \left( \frac{\sigma_\epsilon^2}{2} \right)^2 \\
&\quad \left[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{\text{ap}}^6 \right\rangle + 9 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^4 \right\rangle \right. \\
&\quad \left. + 18 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^2 \right\rangle + 6 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \left\langle \mathcal{M}_{s,2}^3 \right\rangle \right] + 600 \mathcal{M}_{g,2}^3 \left( \frac{\sigma_\epsilon^2}{2} \right)^3 \\
&\quad \left[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} \left\langle \mathcal{M}_{\text{ap}}^4 \right\rangle + 4 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^2 \right\rangle + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \left\langle \mathcal{M}_{s,2}^2 \right\rangle \right] \\
&+ 600 \mathcal{M}_{g,2}^4 \left( \frac{\sigma_\epsilon^2}{2} \right)^4 \left[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} \left\langle \mathcal{M}_{\text{ap}}^2 \right\rangle + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \left\langle \mathcal{M}_{s,2} \right\rangle \right] \\
&+ 120 \mathcal{M}_{g,2}^5 \left( \frac{\sigma_\epsilon^2}{2} \right)^5 \left. \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 \right\} - \left\langle \mathcal{M}_{\text{ap}}^5 \right\rangle^2
\end{aligned}
$$

$$
\sigma^2 \left[ \widehat{M_{\text{ap}}^6} \right] = \frac{1}{\left( \sum_{\neq} w_{i_1} w_{i_2} w_{i_3} w_{i_4} w_{i_5} w_{i_6} \right)^2} \left\{ \sum_{\neq} w_{i_1} w_{j_1} w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^{12} \right\rangle \right.
$$

$$
\begin{aligned}
&+ 36 \sum_{\neq} w_{i_1}^2 w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^{10} \right\rangle + 450 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^{8} \right\rangle \\
&+ 2400 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\text{ap}}^{6} \right\rangle + 5400 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^4 \mathcal{M}_{\text{ap}}^{4} \right\rangle \\
&+ 4320 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^5 \mathcal{M}_{\text{ap}}^{2} \right\rangle + 720 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^6 \right\rangle + 36 \mathcal{M}_{g,2} \left( \frac{\sigma_\epsilon^2}{2} \right) \\
&\quad \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2} w_{j_2} w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^{10} \right\rangle + 25 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^{8} \right\rangle \\
&\quad + 200 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^{6} \right\rangle + 600 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\text{ap}}^{4} \right\rangle \\
&\quad + 600 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^4 \mathcal{M}_{\text{ap}}^{2} \right\rangle + 120 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^5 \right\rangle \Bigg] + 450 \mathcal{M}_{g,2}^2 \left( \frac{\sigma_\epsilon^2}{2} \right)^2 \\
&\quad \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^{8} \right\rangle + 16 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^{6} \right\rangle \\
&\quad + 72 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^{4} \right\rangle + 96 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^3 \mathcal{M}_{\text{ap}}^{2} \right\rangle \\
&\quad + 24 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^4 \right\rangle \Bigg] + 2400 \mathcal{M}_{g,2}^3 \left( \frac{\sigma_\epsilon^2}{2} \right)^3 \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4} w_{j_4} w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^{6} \right\rangle \\
&\quad + 9 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^{4} \right\rangle + 18 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\text{ap}}^{2} \right\rangle
\end{aligned}
$$

$$+ 6 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^3 \right\rangle \Bigg] + 5400 \mathcal{M}_{g,2}^4 \left( \frac{\sigma_\epsilon^2}{2} \right)^4 \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5} w_{j_5} w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^4 \right\rangle$$

$$+ 4 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{s,2} \mathcal{M}_{\text{ap}}^2 \right\rangle + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2}^2 \right\rangle \Bigg] + 4320 \mathcal{M}_{g,2}^5 \left( \frac{\sigma_\epsilon^2}{2} \right)^5$$

$$\left[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6} w_{j_6} \left\langle \mathcal{M}_{\text{ap}}^2 \right\rangle + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \left\langle \mathcal{M}_{s,2} \right\rangle \right] + 720 \mathcal{M}_{g,2}^6 \left( \frac{\sigma_\epsilon^2}{2} \right)^6 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 w_{i_4}^2 w_{i_5}^2 w_{i_6}^2 \Bigg\}$$

$$- \left\langle \mathcal{M}_{\text{ap}}^6 \right\rangle^2$$

## 3.D Variance of the direct estimator for the aperture mass skewness

### 3.D.1 Notation

Let us begin this section by defining some useful notation. Unless otherwise specified, for an $n$th order computation we assume apertures with $N_g > n$ galaxies within them.

$$M_{s,n} \equiv \frac{\sum_{i=1}^{N_g}(w_i Q_i \epsilon_{t,i})^n}{\left(\sum_{i=1}^{N_g} w_i\right)^n} \equiv \frac{\sum_{i=1}^{N_g}(w_i x_i)^n}{\left(\sum_{i=1}^{N_g} w_i\right)^n} \quad ; \quad x_i \equiv Q_i \epsilon_{t,i} \ ; \tag{3.71}$$

$$M_{g,n} \equiv \frac{\sum_{i=1}^{N_g}(w_i Q_i)^n}{\left(\sum_{i=1}^{N_g} w_i\right)^n} \ . \tag{3.72}$$

From now on we assume all sums with no explicit upper limit to run up to $N_g$. As the summation indices do become rather messy, we shall also define the following simplifying shorthands:

$$\sum_{\neq} \equiv \sum_{i_1} \sum_{i_2 \neq i_1} \cdots \sum_{i_n \neq i_{n-1} \neq \cdots \neq i_1} \ ; \tag{3.73}$$

$$\sum_{\substack{\neq \\ i_a = i_b}} \equiv \sum_{i_1} \sum_{i_2 \neq i_1} \cdots \sum_{i_a \neq \cdots \neq i_1} \cdots \sum_{i_{b-1} \neq \cdots \neq i_1} \sum_{i_{b+1} \neq \cdots \neq i_1} \cdots \sum_{i_n \neq \cdots \neq i_1} \ . \tag{3.74}$$

### 3.D.2 Computation

With this background notation in hand, the variance of $\langle \mathcal{M}_{\rm ap}^3 \rangle$ can be written as

$$\sigma^2\left[\widehat{M_{\rm ap}^3}\right] = \left\langle \left(\pi\theta^2\right)^3 \frac{\sum_{\neq} w_{i_1} w_{i_2} w_{i_3} x_{i_1} x_{i_2} x_{i_3}}{\sum_{\neq} w_{i_1} w_{i_2} w_{i_3}} \cdot \left(\pi\theta^2\right)^3 \frac{\sum_{\neq} w_{j_1} w_{j_2} w_{j_3} x_{j_1} x_{j_2} x_{j_3}}{\sum_{\neq} w_{j_1} w_{j_2} w_{j_3}} \right\rangle - \langle \mathcal{M}_{\rm ap}^3 \rangle^2 \tag{3.75}$$

We proceed as always by averaging over the source galaxies. For the third order variance we then expect four structurally identical terms each corresponding to various permutations of contractions. The prefactor can be found by considering the following scheme. We represent the two groups of indices in a similar shape to a six point correlator and count the number of different contractions that contract[14] an index of the $i$ set with one of the $j$ set. Let us do an example to count all double contractions (see illustration below): For a single contraction we have 9 possibilities, whereas for two contractions we can effectively do all single contractions (9 terms) and delete the contracted indices, leaving us with just four remaining indices. Connecting those gives 4 more possibilities. Finally we divide by the factorial of the number of contractions, as those are interchangeable.

$$\left| \overbrace{i_1\ \overbrace{i_2\ i_3}\ |\ j_1\ \overbrace{j_2\ j_3}} \right| = \frac{9}{2!} \times \left| \overbrace{i_2\ i_3}\ |\ \overbrace{j_2\ j_3} \right| = \frac{9 \cdot 4}{2!} = 18$$

Generalizing to $a$ contractions for two $n$th order index sets this gives $C(n,a) = \frac{n^2 (n-1)^2 \cdots (n-(a-1))^2}{a!}$.
Now we find for the source galaxy averaging

$$\begin{aligned}
A(\epsilon_{t,i_1}\epsilon_{t,i_2}\epsilon_{t,i_3}\epsilon_{t,j_1}\epsilon_{t,j_2}\epsilon_{t,j_3}) &= \gamma_{t,i_1}\gamma_{t,i_2}\gamma_{t,i_3}\gamma_{t,j_1}\gamma_{t,j_2}\gamma_{t,j_3} && \text{(1)}\\
\overbrace{i_1\ i_2\ i_3}\ |\ \overbrace{j_1\ j_2\ j_3} \quad &+ \left(\frac{\sigma_\epsilon^2}{2}\right)\left(Q_{i_1}^2 \gamma_{t,i_2}\gamma_{t,i_3}\gamma_{t,j_2}\gamma_{t,j_3}\delta^K_{i_1,j_1} + 8 \text{ perm.}\right) && \text{(2)}\\
\overbrace{i_1\ \overbrace{i_2\ i_3}}\ |\ \overbrace{j_1\ j_2}\ j_3 \quad &+ \left(\frac{\sigma_\epsilon^2}{2}\right)^2\left(Q_{i_1}^2 Q_{i_2}^2 \gamma_{t,i_3}\gamma_{t,j_3}\delta^K_{i_1,j_1}\delta^K_{i_2,j_2} + 17 \text{ perm.}\right) && \text{(3)}\\
\overbrace{i_1\ \overbrace{i_2\ i_3}}\ |\ \underbrace{j_1\ j_2\ j_3} \quad &+ \left(\frac{\sigma_\epsilon^2}{2}\right)^3\left(Q_{i_1}^2 Q_{i_2}^2 Q_{i_3}^2 \delta^K_{i_1,j_1}\delta^K_{i_2,j_2}\delta^K_{i_3,j_3} + 5 \text{ perm.}\right) && \text{(4)}
\end{aligned} \tag{3.76}$$

We now perform the positional averaging over those terms (1) − (4) individually. In order to shorten similar calculations we note the following identity for the position average corresponding to an $m$ point contraction of an $n$th order variance:

$$\left\langle P\left( \sum_{\substack{\neq \\ i_1=j_1 \\ \cdots \\ i_m=j_m}} w_{i_1}\cdots w_{i_n} w_{j_1}\cdots w_{j_n}\ x_{i_1}\cdots x_{i_n} x_{j_1}\cdots x_{j_n} \right)\right\rangle$$

[14] In this note contraction of two indices means that they are set equal to each other.

$$= \sum_{\neq} w_{i_1}^2 \cdots w_{i_m}^2 w_{i_{m+1}} \cdots w_{i_n} w_{j_{m+1}} \cdots w_{j_n} \langle \mathcal{M}_{s,2}^m \mathcal{M}_{\mathrm{ap}}^{2(n-m)} \rangle \tag{3.77}$$

For the term (1) the index structure in the summation symbol has not changed at all, so we can simply recycle the reasoning to get to the ellipticity averaging calculation. Also adding in the ensemble average we get

$$\begin{aligned}
&\left\langle P \left( \sum_{\neq} w_{i_1} w_{i_2} w_{i_3} x_{i_1} x_{i_2} x_{i_3} \sum_{\neq} w_{j_1} w_{j_2} w_{j_3} x_{j_1} x_{j_2} x_{j_3} \right) \right\rangle \\
&\quad = \sum_{\neq} w_{i_1} w_{i_2} w_{i_3} w_{j_1} w_{j_2} w_{j_3} \langle \mathcal{M}_{\mathrm{ap}}^6 \rangle + 9 \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_2} w_{j_3} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\mathrm{ap}}^4 \rangle \\
&\quad + 18 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} \langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\mathrm{ap}}^2 \rangle + 6 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \langle \mathcal{M}_{s,2}^3 \rangle
\end{aligned} \tag{3.78}$$

Note that in this case we can pull out the prefactor from the summation symbols as the ensemble average quantities are theory values.

For the second set of terms (2) we have one Kronecker delta in place such that we can only contract over the remaining five indices. For example, the first term with the matching weights yields[15]:

$$\begin{aligned}
&\left\langle P \left( \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{j_2 \neq i_1} \sum_{j_3 \neq j_2 \neq i_1} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_2} w_{j_3} Q_{i_1}^2 x_{i_2} x_{i_3} x_{j_2} x_{j_3} \right) \right\rangle \\
&\quad = M_{g,2} \left( \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_2} w_{j_3} \langle \mathcal{M}_{\mathrm{ap}}^4 \rangle + 4 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\mathrm{ap}}^2 \rangle \right. \\
&\qquad \left. + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^3 \langle \mathcal{M}_{s,2}^2 \rangle \right)
\end{aligned} \tag{3.79}$$

All the other permutations simply shift the squares in one of the $w_i$s around, but does not change the result - hence we can simply multiply by 9.

Continuing with the terms in (3) the two Kronecker deltas force us to do either one or no contraction. For the first term the result looks like:

$$\begin{aligned}
&\left\langle P \left( \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{j_3 \neq i_2 \neq i_1} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} Q_{i_1}^2 Q_{i_2}^2 x_{i_3} x_{j_3} \right) \right\rangle \\
&\quad = M_{g,2}^2 \left( \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} \langle \mathcal{M}_{\mathrm{ap}}^2 \rangle + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \langle \mathcal{M}_{s,2} \rangle \right)
\end{aligned} \tag{3.80}$$

Again, all the other permutations yield the same result, so we can multiply by 18.

For the final (4) term no further contractions can be done and, again, all permutations give equivalent answers. For the first term we have

$$\left\langle P \left( \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 Q_{i_1}^2 Q_{i_2}^2 Q_{i_3}^2 \right) \right\rangle = M_{g,2}^3 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^3 \tag{3.81}$$

[15] For an explicit computation of this term, see Appendix 3.D.3

Collecting together all the terms we find the weighted variance of the $\langle \mathcal{M}_{\rm ap}^3 \rangle$ to be

$$
\begin{aligned}
\sigma^2\left[\widehat{M_{\rm ap}^3}\right] = \frac{1}{\left(\sum_{\neq} w_{i_1} w_{i_2} w_{i_3}\right)^2} \Bigg\{ & \\
& \sum_{\neq} w_{i_1} w_{i_2} w_{i_3} w_{j_1} w_{j_2} w_{j_3} \langle \mathcal{M}_{\rm ap}^6 \rangle + 9 \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} w_{j_2} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\rm ap}^4 \rangle \\
& + 18 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_1} \langle \mathcal{M}_{s,2}^2 \mathcal{M}_{\rm ap}^2 \rangle + 6 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \langle \mathcal{M}_{s,2}^3 \rangle \\
& + 9 M_{g,2} \left(\frac{\sigma_\epsilon^2}{2}\right) \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} w_{j_2} \langle \mathcal{M}_{\rm ap}^4 \rangle + 4 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_1} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\rm ap}^2 \rangle \\
& \qquad + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^3 \langle \mathcal{M}_{s,2}^2 \rangle \Bigg] \\
& + 18 M_{g,2}^2 \left(\frac{\sigma_\epsilon^2}{2}\right)^2 \Bigg[ \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_1} \langle \mathcal{M}_{\rm ap}^2 \rangle + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \langle \mathcal{M}_{s,2} \rangle \Bigg] \\
& + 6 M_{g,2}^3 \left(\frac{\sigma_\epsilon^2}{2}\right)^3 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \Bigg\} - \langle \mathcal{M}_{\rm ap}^3 \rangle^2
\end{aligned}
\tag{3.82}
$$

### 3.D.3 Explicit computation of one third order term

We now compute one contraction term explicitly and show that all permutations and higher order contractions can be computed in a similar fashion. As a first step let us write down all the possibilities that the six indices can take. In here the $a$th element of each tuple shows which index the $j_a$ is, it is either $i_1$, $i_2$, $i_3$, or none of those which is labelled $\neq$. The horizontal lines separate sets of tuples which have the same number of unequal indices. Note that counting through the tuples we get the numbers from the contractions.

$$(\neq \neq \neq)$$

---

$$(i_1 \neq \neq) \quad (\neq i_1 \neq) \quad (\neq \neq i_1)$$
$$(i_2 \neq \neq) \quad (\neq i_2 \neq) \quad (\neq \neq i_2)$$
$$(i_3 \neq \neq) \quad (\neq i_3 \neq) \quad (\neq \neq i_3)$$

---

$$(i_1\, i_2 \neq) \quad (i_1 \neq i_2) \quad (\neq i_1\, i_2)$$
$$(i_1\, i_3 \neq) \quad (i_1 \neq i_3) \quad (\neq i_1\, i_3)$$
$$(i_2\, i_3 \neq) \quad (i_2 \neq i_3) \quad (\neq i_2\, i_3)$$
$$(i_2\, i_1 \neq) \quad (i_2 \neq i_1) \quad (\neq i_2\, i_1)$$
$$(i_3\, i_1 \neq) \quad (i_3 \neq i_1) \quad (\neq i_3\, i_1)$$
$$(i_3\, i_2 \neq) \quad (i_3 \neq i_2) \quad (\neq i_3\, i_2)$$

---

$$(i_1\, i_2\, i_3) \quad (i_2\, i_1\, i_3) \quad (i_3\, i_1\, i_2)$$
$$(i_1\, i_3\, i_2) \quad (i_2\, i_3\, i_1) \quad (i_3\, i_2\, i_1)$$

The term we deal with is the first one in (1) where $j_1$ is set equal to $i_1$. In a first step we rewrite the six summation symbols in terms of summations that solely consist of unequal indices. For this we can only choose the tuples that have $i_1$ as a first entry. We then find successively

$$\begin{aligned}
&\sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{j_1} \sum_{j_2 \neq j_1} \sum_{j_3 \neq j_2 \neq j_1} w_{i_1} w_{i_2} w_{i_3} w_{j_1} w_{j_2} w_{j_3} Q_{i_1} x_{i_2} x_{i_3} Q_{j_1} x_{j_2} x_{j_3} \delta^K_{i_1,j_1} \\
&\qquad = \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{j_2 \neq i_1} \sum_{j_3 \neq j_2 \neq i_1} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_2} w_{j_3} Q_{i_1}^2 x_{i_2} x_{i_3} x_{j_2} x_{j_3} \\
&\qquad = \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_2} w_{j_3} Q_{i_1}^2 x_{i_2} x_{i_3} x_{j_2} x_{j_3} \\
&\qquad + \sum_{\neq} \left( w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_3} Q_{i_1}^2 x_{i_2}^2 x_{i_3} x_{j_3} + w_{i_1}^2 w_{i_2} w_{i_3}^2 w_{j_3} Q_{i_1}^2 x_{i_2} x_{i_3}^2 x_{j_3} \right) \\
&\qquad + \sum_{\neq} \left( w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_2} Q_{i_1}^2 x_{i_2}^2 x_{i_3} x_{j_2} + w_{i_1}^2 w_{i_2} w_{i_3}^2 w_{j_2} Q_{i_1}^2 x_{i_2} x_{i_3}^2 x_{j_2} \right) \\
&\qquad + \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \left( Q_{i_1}^2 x_{i_2}^2 x_{i_3}^2 + Q_{i_1}^2 x_{i_3}^2 x_{i_2}^2 \right)
\end{aligned}$$

$$= \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} w_{j_2} Q_{i_1}^2 x_{i_2} x_{i_3} x_{j_1} x_{j_2} + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 \; Q_{i_1}^2 x_{i_2}^2 x_{i_3}^2$$
$$+ 2 \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} Q_{i_1}^2 x_{i_2} x_{i_3} x_{j_1} \left(w_{i_2} x_{i_2} + w_{i_3} x_{i_3}\right)$$

where in the first step we applied the delta, in the second one subbed in all the relevant terms and in the third one renamed indices and combined equal terms. Note that the number of tuples chosen for each number of $\neq$ symbols does match the one from the contraction formalism. Now we apply the position and ensemble averaging.

$$\left\langle P \left( \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{j_1} \sum_{j_2 \neq j_1} \sum_{j_3 \neq j_2 \neq j_1} w_{i_1} w_{i_2} w_{i_3} w_{j_1} w_{j_2} w_{j_3} Q_{i_1} x_{i_2} x_{i_3} Q_{j_1} x_{j_2} x_{j_3} \delta^K_{i_1, j_1} \right) \right\rangle$$
$$= \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} w_{j_2} M_{g,2} \langle \mathcal{M}_{\mathrm{ap}}^4 \rangle + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 M_{g,2} \langle \mathcal{M}_{s,2}^2 \rangle$$
$$+ 2 \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} \left(w_{i_2} + w_{i_3}\right) M_{g,2} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\mathrm{ap}}^2 \rangle$$
$$= \sum_{\neq} w_{i_1}^2 w_{i_2} w_{i_3} w_{j_1} w_{j_2} M_{g,2} \langle \mathcal{M}_{\mathrm{ap}}^4 \rangle + 2 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3}^2 M_{g,2} \langle \mathcal{M}_{s,2}^2 \rangle$$
$$+ 4 \sum_{\neq} w_{i_1}^2 w_{i_2}^2 w_{i_3} w_{j_1} M_{g,2} \langle \mathcal{M}_{s,2}^1 \mathcal{M}_{\mathrm{ap}}^2 \rangle$$

where in the last step we noted the argument of the sum with brackets is symmetric, and hence the results for both terms are equal. This is exactly the result we would have expected from the contractions on the subset excluding $i_1$ and $j_1$.
Looking at the other eight permutations, the only difference is that we choose different indices at the start - however there will always be equally many and all the steps are essentially mirrored, therefore we can just multiply the result we got by 9.

## 3.E Distribution of the statistics

In this appendix we show the sample distribution of the various aperture mass statistics within the 819 SLICS mocks in Fig. 3.11. In order to being able to compare the shapes of the distributions for various aperture radii we shift and normalise each pdf such that it has zeros mean and unit variance. Let us at first focus on the measurements on the ensemble for which the shape noise has been set to zero. We see that practically only the first order measurement is well approximated by a Gaussian sampling distribution across all aperture scales. Going to higher orders one observes a clear deviation from a normal sampling distribution where the difference increases for higher orders and larger aperture radii. When shape noise is included the curves look similar for large aperture radii, while the measurements for small scales are dominated by shape noise

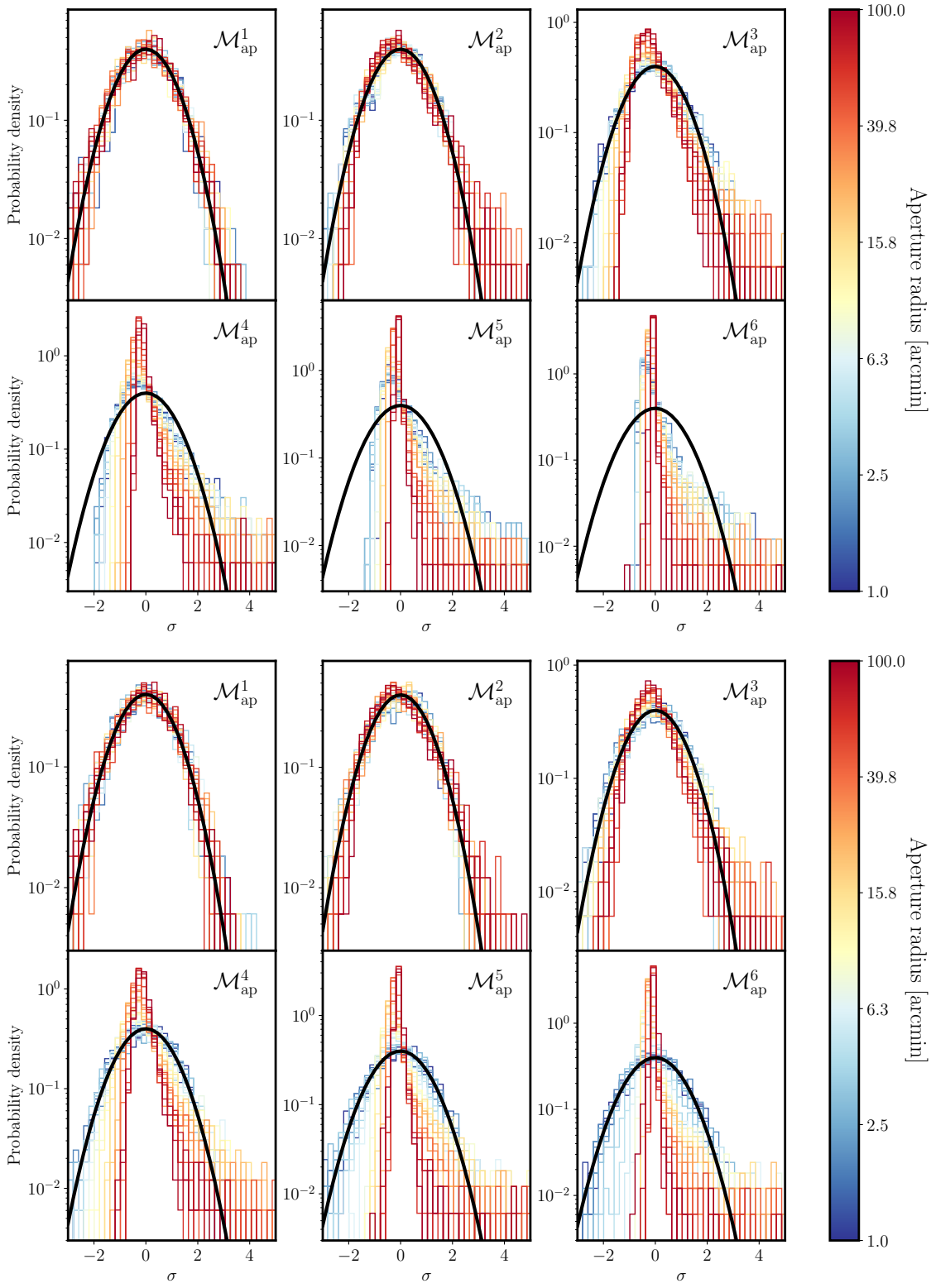


Figure 3.11 Normalised sample distribution of the aperture mass moments for the first six orders of the statistics in the SLICS mocks. In the upper panel we show the results for the noiseless case while the lower panel displays the histograms when including the shape noise contribution. The black curve indicates a standard normal distribution.

and hence they do converge to the normal distribution again. Looking at the overall difference to the Gaussian approximation one can say that up until second order the approximation is roughly valid while for higher orders one expects corrections that need to be thoroughly modeled. We note that these results might change when running the estimators on a larger survey: As each estimate is given as a sum over estimates stemming from individual apertures and as the number of apertures is proportional to the survey area, the CLT will make the sampling distribution of the estimator converge to a Gaussian. One might already see a hint of this feature in the upper panel where the sampling distribution of small aperture radii is closer to a normal distribution than the one for larger radii. As the number of individual apertures per estimate is also proportional to $R_{\mathrm{ap}}^{-2}$ and no shape noise was present in this case, we might interpret this feature as a consequence of the CLT.

# The Information Content of Projected Galaxy Fields

4

## 4.1 Motivation

On a high level, the potential constraining power of cosmological weak lensing surveys depends on the observed area and on the survey depth. As a theoretical description of the nonlinear density field is lacking, the full information cannot be retrieved and instead one performs cosmological analyses with the help of summary statistics. The most widely used statistic is the power spectrum containing only the Gaussian contribution of the available information. In the context of three dimensional fields there have been proposed linearization methods (Neyrinck et al., 2009; Joachimi et al., 2011; Neyrinck et al., 2011) that successfully remap some of the nonlinear information into the second order statistics; however the information gain is strongly depleted once shot noise is taken into account. A more complete way to extract cosmological information from spectroscopic surveys can be formulated in terms of forward modelling approaches that do not need to compress the observed data. Recently, the proposed models in (Kitaura & Enßlin, 2008; Jasche & Kitaura, 2010; Jasche & Wandelt, 2013a,b; Wang et al., 2014) have been successfully applied to N-body simulations and real galaxy redshift surveys (Lavaux & Jasche, 2016; Leclercq et al., 2017) to reveal a wealth of information of the dark matter density field and its phase space distribution in our nearby Universe. Due to the high dimensionality of the resulting posteriors, those models have mainly been used to reconstruct the three dimensional matter distribution when using a fixed cosmology.

In this chapter we aim test how well an idealistic forward model performs for projected tracer fields with a line of sight resolution that might be acquired from imaging surveys. To assess the information contained in the resulting posterior we do not keep the cosmology fixed and allow the amplitude of the transformed power spectrum to vary. The information content can

then be rephrased as the the signal-to-noise of the estimated amplitude parameter after having marginalized over the projected mass field itself.

This chapter is organised as follows: In §4.2 we introduce the hierarchical model adopted for our reconstruction and test the validity of our parametrizations on the Quijote simulation suite. In §4.3 we give an overview of the Hamiltonian Monte Carlo sampling algorithm and lay out some specific choices we made for our implementation. In §4.4 we first validate our model on a suite of lognormal simulations, then apply it to the Quijote suite and finally compare the reconstruction confidence of the hierarchical model to the expected signal-to-noise using standard analysis methods.

## 4.2 Model

For this work we suppose that we are given a set of galaxies on a two dimensional plane. We then cover the plane with a regular grid and assign the galaxies to the grid cells, yielding a set of galaxy counts $N \equiv \{N_c\}$. A hierarchical model connecting the observations to a Gaussian mass field $\delta^{\mathrm{lin}}$ with a power spectrum $P$ parametrized by a set of parameters $\Pi_P$ can then be schematically written down as:

$$\mathbb{P}\left(\delta^{\mathrm{lin}}, \Pi_P \,\middle|\, N, B, P\right) \propto \mathbb{P}\left(N \,\middle|\, \delta^{\mathrm{nl}}, B\right) \mathbb{P}\left(\delta^{\mathrm{nl}} \,\middle|\, \delta^{\mathrm{lin}}\right) \mathbb{P}\left(\delta^{\mathrm{lin}}, \Pi_P \,\middle|\, P\right) \mathbb{P}(\Pi_P) \,, \tag{4.1}$$

where $\delta^{\mathrm{nl}}$ denotes a nonlinear mass field for which the galaxies are assumed to be biased tracers. The model for galaxy bias is encoded in the bias operator $B$. In the following we will give our modelling choices for each contribution to Eq. (4.1).

### The Poissonian likelihood

The first term models the way on how the set of discrete tracers is sampled onto the dark matter fluid; we assume this to follow an inhomogeneous Poisson process (Layzer, 1956; Peebles, 1980):

$$\mathbb{P}\left(N \,\middle|\, \delta^{\mathrm{nl}}, B, \bar{n}\right) = \mathbb{P}\left(\bar{n}\right) \prod_c \mathrm{e}^{-n_c} \frac{n_c^{N_c}}{N_c!} \,, \tag{4.2}$$

$$n_c \equiv \bar{n}\left(1 + B\left[\delta^{\mathrm{nl}}\right]_c\right) \,, \tag{4.3}$$

where we introduced the galaxy density field $n$ which depends on the mean number density of observed galaxies $\bar{n}$ and the physical connection $B$ between the galaxies and dark matter. We do furthermore put a logarithmic prior on $\bar{n}$, such that one can perform an analytic marginalization:

$$\mathbb{P}\left(N \,\middle|\, \delta^{\mathrm{nl}}, B\right) \propto \prod_c \left(1 + B\left[\delta^{\mathrm{nl}}\right]_c\right)^{N_c} \left(\sum_c 1 + B\left[\delta^{\mathrm{nl}}\right]_c\right)^{-N_{\mathrm{tot}}} . \tag{4.4}$$

In the following we will make the simplifying assumption of $B \equiv 1$.

The second term in Eq. (4.1) corresponds to a physical model that evolves a linear density field to a nonlinear one. For this work we will assume a local model of structure formation, such that

$$\mathbb{P}\left(\delta^{\mathrm{nl}} \,\middle|\, \delta^{\mathrm{lin}}\right) = \prod_c \delta^D\left(\delta_c^{\mathrm{nl}} - G^{-1}\left(\delta_c^{\mathrm{lin}}\right)\right) \,, \tag{4.5}$$

where $G$ is a function that aims at inverting structure formation. For this work we will consider two different forms of $G^{-1}$. The first is a logarithmic transformation, while the second one is Schechter-like:

$$\begin{aligned} &\text{Logarithmic}: \\ &G^{-1}(\delta^{\mathrm{lin}}) = \mathrm{e}^{\delta^{\mathrm{lin}} - \sigma^2/2} - 1 \,; \end{aligned} \tag{4.6}$$

$$\begin{aligned} &\text{Schechter}: \\ &G^{-1}(\delta^{\mathrm{lin}}) = n\, \mathrm{e}^{a_1\delta^{\mathrm{lin}} - a_1^2\sigma^2/2}\left(1 + \mathrm{e}^{(\delta^{\mathrm{lin}} - \delta_0)t}\right)^{\frac{a_2 - a_1}{t}} - 1 \,, \end{aligned} \tag{4.7}$$

where $\sigma^2$ is the variance of the linear overdensity field and the normalization function $n$ is defined to yield $\langle \delta^{\mathrm{nl}} \rangle \equiv 0$. Note that the Schechter-like transformation is constructed to interpolate between two biased logarithmic transformations around a characteristic scale $\delta_0$ with an interpolation 'width' described by $1/t$.

### The Gaussian prior

Assuming that the function $G$ completely gaussianizes the $\delta^{\mathrm{nl}}$, the linear field will be fully described by its correlation function $\xi$, such that we can write down a corresponding prior as

$$\mathbb{P}\left(\delta^{\mathrm{lin}}, \Pi_P \,\middle|\, \xi\right) = \frac{1}{\sqrt{(2\pi)^{n_{\mathrm{pix}}}|\xi|}} \exp\left[-\frac{1}{2}\sum_{c,c'} \delta_c^{\mathrm{lin}} \xi_{cc'}^{-1} \delta_{c'}^{\mathrm{lin}}\right] \,, \tag{4.8}$$

where $n_{\mathrm{pix}}$ denotes the total number of pixels in the grid and $|\xi|$ is the determinant of the correlation matrix. In order to circumvent the computationally not feasible matrix inversion and products in the exponent we evaluate Eq. (4.8) in its harmonic basis:

$$\mathbb{P}\left(\delta^{\mathrm{lin}}, \Pi_P \,\middle|\, P\right) \propto \left(\prod_c P_c^{-1/2}\right) \exp\left[-\frac{1}{2}\sum_c \delta_c^{\mathrm{lin}}\, \mathcal{F}^{-1}\left[\frac{\widetilde{\delta^{\mathrm{lin}}}}{P}\right]_c\right] . \tag{4.9}$$

For the remainder of this work we will assume that also $P$ is parametrizable by a function that interpolates between two different power laws:

$$P(k; A, k_0, a_1, a_2, s) = A\left(\left(\frac{k}{k_0}\right)^{a_1 s} + \left(\frac{k}{k_0}\right)^{a_2 s}\right)^{-\frac{1}{s}} \,. \tag{4.10}$$

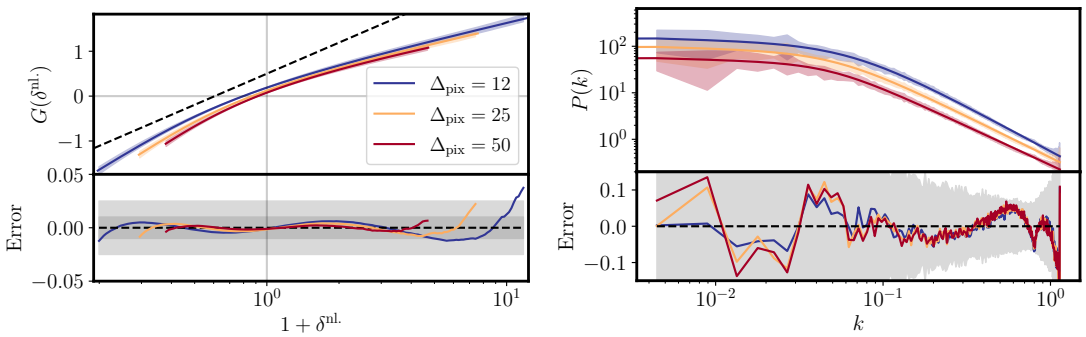


Figure 4.1 *Left hand side*: Gaussianizing transformations for an angular resolution of $\approx 4\,\mathrm{Mpc}/h$ and a set of projection depths in units of the resolution $\Delta_{\rm pix}$. The shaded region in the upper panel corresponds the the standard deviation of the numerical transformation equations across the ensemble. The solid lines in the upper panel correspond to the best fit model Eq. (4.7). The black dashed line indicates the expected scaling from a logarithmic transformation model Eq. (4.6). The lower panel displays the relative error between the best fit model and the mean numerical transformation. The shaded regions in the lower panel display the $2.5\%$ and $1\%$ errorbands. *Right hand side*: Same as the plot on the left, but for the power spectra of the transformed fields. In the lower panel we plot the $5\%$ errorband as well as the statistical spread of measurements across the ensemble.

### Accuracy of parametrizations

For assessing the applicability of the parametrizations Eq. (4.7) and Eq. (4.10) to N-body simulations we test their accuracy with help of the Quijote suite (Villaescusa-Navarro et al., 2020). In particular, we make use of the ensemble of 100 high resolution simulations in which $1024^3$ particles were evolved within a $1h^{-1}\mathrm{Gpc}$ box. After retrieving the snapshots at $z = 1$ we assign the particles onto a regular mesh consisting of $256^3$ voxels using a NGP scheme. From those we create projected mass fields by specifying a projection depth and summing up the content in the corresponding voxels.

For each mass slab $i$ we then employ inverse transform sampling to construct a linearising function $G_i^{\rm num}$ that maps the projected overdensity field to a field following a Gaussian distribution with zero mean and a variance matching the one we would have gotten when log-transforming the overdensity field. Averaging over all the $G_i^{\rm num}$ defines our numeric gaussianizing function $G^{\rm num}$ to which we fit the Schechter transformation Eq. (4.7). The best fit parameters then determine our model for $G^{-1}$. The results from this procedure for our chosen angular resolution of $\approx\ 4\ \mathrm{Mpc}/h$ are shown in the left panel of Fig. 4.1 and we see that $G$ does give a percent accurate fit for nearly all pixels. We furthermore note that while $G$ scales similarly to a

logarithmic model around the mode of $\delta^{\rm nl}$, it does quite strongly deviate from such a model for moderately overdense and underdense regions.

After having fixed the linearization procedure we transform each of the projected overdensity fields according to $G$ and compute the associated power spectra $P_i$. Again, we determine the best fit parameters of the model Eq. (4.10) against the mean of the $P_i$ to define our final model for $P$. We show the numeric result in the right hand side of Fig. 4.1. As for the transformations we find a reasonable agreement across all scales that is well within the statistical spread of the true spectra. We repeated the analysis described above for different meshs consisting of $64^2,\ 128^2$ and $512^2$ pixels, as well as for power spectrum fits to models where the logarithmic transformation Eq. (4.6) had been applied; none of them gave results to which our discussion does not apply.

## 4.3 Sampling method

### 4.3.1 Hamiltonian Monte Carlo Sampling

We employ a Hamiltonian Monte Carlo (HMC) scheme (Duane et al., 1987) to efficiently sample from the high dimensional distribution Eq. (4.1). This method evades the curse of dimensionality by exploring level sets of a distribution $\mathcal{P} \propto \mathrm{e}^{-\mathcal{H}}$ in which the Hamiltonian $\mathcal{H}$ is defined as

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) \equiv \frac{1}{2}\boldsymbol{p}^T \mathbf{M}^{-1} \boldsymbol{p} + \psi(\boldsymbol{q}) \; ; \tag{4.11}$$

$$\psi(\boldsymbol{q}) \equiv -\ln \mathbb{P}(\boldsymbol{q}) \; , \tag{4.12}$$

where we assume the auxiliary momentum variables $\boldsymbol{p}$ to follow a Gaussian distribution, $\boldsymbol{p} \sim \mathcal{G}\,(0, \mathbf{M})$. From this formulation one can construct a valid Markov chain for the original posterior $\mathbb{P}$ by marginalizing over the momenta and for each drawn sample of $\boldsymbol{p}$ evolve the system to a new location $(\boldsymbol{q}', \boldsymbol{p}')$ in phase space according to the Hamilton equations of motion

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} = \mathbf{M}^{-1}\boldsymbol{p} \tag{4.13}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} = -\nabla_{\boldsymbol{q}}\psi(\boldsymbol{q}) \; . \tag{4.14}$$

Due to numerical inaccuracies the Hamiltonian will not be exactly conserved along the trajectories; in order to still satisfy the detailed balance condition one then needs to invoke a Metropolis-Hastings rejection step prior to updating the chain with the value of $\boldsymbol{q}'$. For more complete reviews of HMC see i.e. Neal (2012).

Adopting the HMC framework to the posterior Eq. (4.1) we have $\boldsymbol{q} = \{\delta^{\rm lin}, \Pi_P\}$ and by making

use of Eqs. (4.9) and (4.4) the potential $\psi$ becomes

$$\psi\left(\boldsymbol{q}\right) = \psi_{\mathrm{Poiss}}\left(\delta^{\mathrm{lin}}\right) + \psi_{\mathrm{Gauss}}\left(\delta^{\mathrm{lin}}, \Pi_P\right) \; ; \tag{4.15}$$

$$\psi_{\mathrm{Poiss}} = N_{\mathrm{tot}} \log\left[\sum_c \left(1 + G^{-1}\left(\delta_c^{\mathrm{lin}}\right)\right)\right] - \sum_c N_c \left(1 + G^{-1}\left(\delta_c^{\mathrm{lin}}\right)\right) \tag{4.16}$$

$$\psi_{\mathrm{Gauss}} = \frac{1}{2}\sum_c \left(\ln(P_c) + \delta_c^{\mathrm{lin}} \mathcal{F}^{-1}\left[\frac{\widetilde{\delta^{\mathrm{lin}}}}{P}\right]_c\right) \; , \tag{4.17}$$

where $N_{\mathrm{tot}}$ denotes the total number of observed tracers. The corresponding expression of the gradient is more cumbersome and we refer to Appendix 4.A for the details.

### 4.3.2 Implementation specifics

As for all sampling schemes there exist multiple knobs that need to be tweaked in order to facilitate an efficient exploration of our posterior. In this subsection we give a top level overview of the choices for our implementation.

For the mass matrix $\mathbf{M}$ we employ a low rank approximation of the Hessian of Eq. (4.1). Our most efficient solution turned out to be formulated in harmonic space and it allows for a drawing of complex momenta that is bound by FFT operations, as well as for a linearly scaling complexity for the matrix vector product in Eq. (4.13). We note that employing a diagonal mass matrix associated with the Hessian of the Gaussian part of the potential did give similarly good convergence properties.

As advocated by the standard literature we invoke a Leapfrog integrator to discretise the evolution equations Eqs. (4.13) and (4.14). This method is expected to be effective as it is a second order symplectic discretization scheme that will conserve the Hamiltonian for a well chosen stepsize. We also checked whether higher order palindromic integrators (see Creutz & Gocksch (1989) for the original formulation, or Hernández-Sánchez et al. (2021) for a first application to cosmology) result in an effective speedup; however for all of our probed grid resolutions the Leapfrog method remained the most efficient integrator.

In order to choose a useful stepsize for the integrator we apply a dual averaging scheme (Nesterov, 2009; Hoffman & Gelman, 2014) during the burnin stage that iteratively adapts the stepsize to a value that will result in some specified acceptance rate $\delta$ during the sampling phase. Once burnin is over we fix the stepsize to its final value $\epsilon$ in the iteration.

Due to the high dimensionality of the problem we also need to worry about the memory footprint of the individual chains. In this work we are mainly concerned with the chains of the power spectrum parameters $\Pi_p$ and therefore we save those as a whole, but do the same only

for a fraction of the latent field parameters $\delta^{\mathrm{lin}}$. As the convergence properties of those variables depend on the number of tracers in the corresponding pixel we make sure that our selection does include sufficiently many overdense and underdense regions.

## 4.4 Results

### 4.4.1 General strategy

For the remainder of this work we will solely concern ourselves with models $\mathcal{M}$ that vary the power spectrum amplitude $A$ jointly with the cosmological field $\delta^{\mathrm{lin}}$ and fix the remaining power spectrum parameters to their best fit values. We furthermore choose the prior value $A^*$ of the amplitude to some value between two thirds and three half times its best fit value and choose this value $A_0$ as the initial value for the amplitude in $\mathcal{M}$. A possible starting position of the latent field, $\delta_0^{\mathrm{lin}}$, can be chosen as a random Gaussian field constructed to match the prior power spectrum with a strongly reduced amplitude.

When running the model $\mathcal{M}$ with these initial conditions we found that the burnin period becomes very prolonged as the starting point is in strong conflict with the Poisson likelihood and a very small stepsize becomes necessary to navigate the chains to their stationary territory[1]. To circumvent most of the complexity we adopt a nested burnin strategy where in a first step we run a simpler model $\mathcal{M}'$ in which we also fix $A \equiv A_0$. Burning in this model with $\delta_0^{\mathrm{lin}}$ is fast and choosing some sample from $\mathcal{M}'$ once stationarity is reached yields a better starting configuration $\delta_0^{\mathrm{lin}}$ for the full model $\mathcal{M}$ that does now burn in much quicker.

Additionally, we note that all our models for generating mock data, as well as the sampling procedure depend on the following set of hyperparameters which are drawn according to some random seeds $r$ for each chain:

(i) The cosmic initial condition $r_{\mathrm{cosmo}}$ that gives rise to to ‘true‘ projected density field.

(ii) The Poisson sampling process $r_{\mathrm{Poiss}}$ that sets the data information in our reconstruction procedure

(iii) The values $r_{\mathrm{ini}}$ used to initialize the reconstruction algorithm

(iv) The values $r_{\mathrm{mom}}$ that specify the order on how to draw the momentum variables during the reconstruction.

[1] We tried various other initialization choices for $\delta_0^{\mathrm{lin}}$ and $A_0$. For each of them we found the same pathological behaviour.

If we want to make a solid prediction about the information content in the power spectrum amplitude $A$ we would formally need to marginalize over a large set of seed configurations. We can get rid of one dimension when making the assumption that $r_{\rm ini}$ and $r_{\rm mom}$ do not influence each other, and we collect both seeds in a new one, $r_{\rm hmc}$. Checking the dispersion of the outcomes when varying over $r_{\rm hmc}$ for fixed $r_{\rm cosmo}$ and $r_{\rm Poiss}$ is then equivalent to assessing the convergence property of the chains, i.e. by virtue of the Gelman-Rubin diagnostics (Gelman & Rubin, 1992). Varying over the remaining two seeds is necessary and we do this for our analysis.

### 4.4.2 Validation on lognormal simulations

While the point transformations $G$ do a good job in removing nonlinearities of the mass field, they are not sufficient to fully gaussianize the field which renders the prior Eq. (4.9) formally wrong. In order to test our implementation we additionally apply our model on an ensemble of lognormal mocks. In particular, we obtain a tracer realization as follows:

(i) We specify a resolution and spatial extent of the slab, as well as a tracer sampling density $\bar{n}$

(ii) We generate a Gaussian random field $g$ having a power spectrum $P_g$ that is tuned to match the best fit power spectrum of the Quijote ensemble of the corresponding slab specifics

(iii) We generate a lognormal field as $\delta^{\rm ln}$ by applying the transformation Eq. (4.6) to $g$

(iv) We treat each pixel as a volume element from the corresponding Quijote slab and Poisson sample tracers into the pixels

In Fig. 4.2 we show as an example the results of a single chain which was run on a $256^2$ grid and we choose to evolve the system for 20 timesteps before updating the chain; we furthermore set $A_0 \equiv 0.7A^*$ and choose $\bar{n} \equiv 0.002/{\rm Mpc}^3$. Looking at the first burnin stage, using the model $\mathcal{M}'$, we see that the latent field has burned in to the prior power spectrum after around 100 iterations. Moving to the full model $\mathcal{M}$ we see that both, the power spectra measured from the latent chains and the theoretical evaluation based on the chain for the amplitude $A$ give results that are consistent with each other and with the true mass field. When tracking the evolution of the amplitude chain we see that during the burnin stage of $\mathcal{M}$ it evolves from the prior value to the true amplitude that was used for generating the mock data[2]. After burnin has finished the chain oscillates around the true value and will therefore provide an unbiased estimate.

[2] We note that for this example we used a less efficient stepsize adaptation algorithm to explicitly showcase the evolution of the amplitude from the prior to the truth.

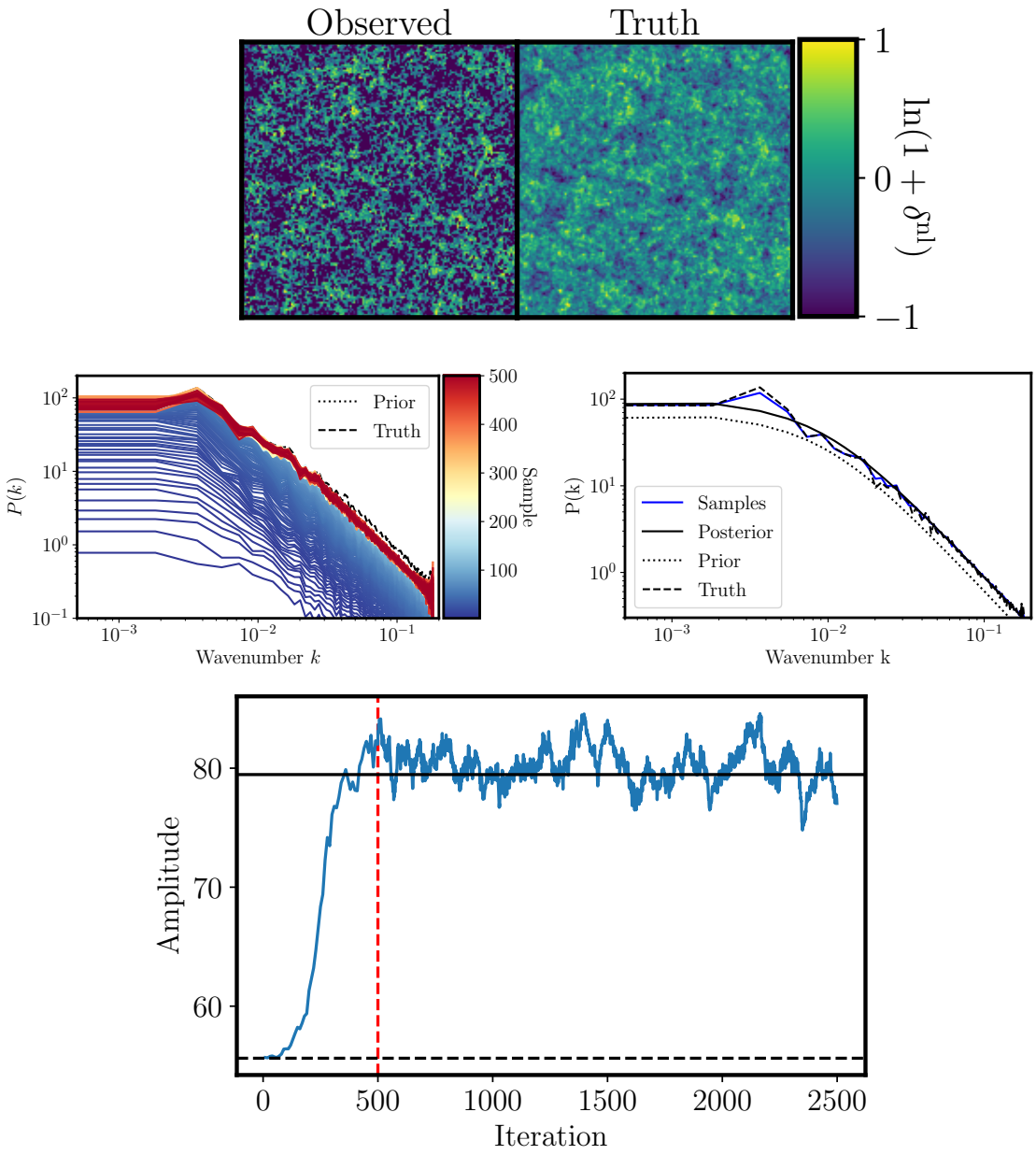


Figure 4.2 Sampling of the posterior Eq. (4.1) on a lognormal mock. *Upper panel:* Comparison of the observed galaxy field to the underlying true field. *Middle left panel:* Power spectra of the proposed latent fields during burning of the simplified model $\mathcal{M}'$. *Middle right panel:* Results of the sampling stage when using the full model $\mathcal{M}$. The blue errorband indicates the variance of the power spectra measured from the proposed latent fields whereas the black solid line gives evaluates the theoretical power spectrum model with the mean of the proposed values for the amplitude. *Lower panel:* The amplitude chain of the model $\mathcal{M}$ during the burnin stage (left of the red dashed line) and the sampling stage. The solid black line indicates the value that was used for generating the mock data while the dashed black line indicates the mean of the prior we did put on $A$.

| Name | Nside | Depth $[\Delta_{\rm pix}]$ | $G^{-1}$ | Snapshots | Poisson realizations | Tracer densities | Initial conditions |
|---|---|---|---|---|---|---|---|
| Schechter_ HighRes | 512 | 50 | Schechter | 5 | 4 | 4 | 6 |
| Schechter_ BaseRes | 256 | 25 | Schechter | 100 | 4 | 4 | 8 |
| Schechter_ LowRes | 128 | 12 | Schechter | 5 | 8 | 4 | 6 |
| Schechter_ vLowRes | 64 | 6 | Schechter | 5 | 8 | 4 | 6 |
| Log_ HighRes | 512 | 50 | Log | 5 | 4 | 4 | 6 |
| Log_ BaseRes | 256 | 25 | Log | 5 | 4 | 4 | 6 |

Table 4.1 Parameter settings for the various ensemble runs performed on snapshots from the Quijote simulation suite. For each of the chains the four tracer densities that are varied over are $0.002,\ 0.005,\ 0.01,\ \text{and}\ 1.0$ tracers per inverse Mpc$^3$. The Schechter_BaseRes run probes the whole Quijote ensemble and consists of the most $(100 \times 4 \times 4 \times 8 = 12800)$ individual chains. Each chain consists of $5000$ Hamiltonian samples.

### 4.4.3 Application to the Quijote simulation suite

#### Overview of ensemble runs

We now turn to the runs on the Quijote ensemble. As discussed in §4.4.1, it is required to vary over cosmological ensembles, as well as over different Poisson sampling realizations. Additionally, we aim to investigate how the reconstruction confidence of our model is affected by the pixel resolution and the sampling density of tracers. In Table 4.1 we summarize the different choices of the hyperparameters for each analysis used in this work.

For each chain we adopt the nested burnin strategy: we sample from model $\mathcal{M}'$ for 500 times before switching to $\mathcal{M}$ which we burn in for another $1000$ steps after which the $5000$ samples that will be used for the subsequent analysis are generated. In order to obtain a new sample we evolve the equations of motion (4.13) for $40$ timesteps. To test the convergence of the chains we furthermore save the full chains for around $5 - 20$ per cent of the latent field pixels, depending on the grid resolution.

### Reconstruction accuracy

Before moving to the results of the ensemble we show in Fig. 4.3 the real space latent field reconstructions of a single chain when using the Schechter transformation Eq. (4.7) or the logormal one, Eq. (4.6). We see that while neither of the transformations succeeds in removing the filamentary structure, the Schechter model does produce less extended tail of overdensities as compared to the lognormal model. We can predict this feature from Fig. 4.1 in which the gaussianizing transformation grows slower than the logarithmic graph and thus will map a larger range of overdensities to an equal sized bin in the transformed field. On the level of an individual sample we see that both models are confident in sampling similar structures in overdense regions while for underdense spots the models resort to their gaussian prior and will therefore not be able to predict the filamentary structure by itself. Finally, for the mean field and its nonlinearized version we again observe a good reconstruction of the high ends of the density field while regions with little data information appear washed out.

Due to the filaments in the latent field and the fact that the Gaussian prior 4.9 does not include higher order clustering statistics, we expect the reconstruction to yield a slightly biased amplitude. In Fig. 4.4 we quantify the magnitude of this effect and its dependence on the galaxy sampling density on the Schechter_BaseRes and Log_BaseRes runs. As a sanity check we also include the lognormal reconstructions LNMock_BaseRes and we find that they give unbiased results. We also see that the Schechter transformation produces a much smaller bias than the lognormal one which we attribute to the fact that it facilitates a gaussianization of the nonlinear field at the one point level: With the tracers only being able to probe the high density tail the logarithmic model implicitly extends this tail to underdense regions, resulting in the power spectrum amplitude to be driven to a higher value. The fewer tracers are observed the more severe this effect and the resulting bias will become. The Schechter model does not suffer from this feature and thus its bias remains nearly constant across all $\bar{n}$. Repeating the analysis for different gridding scales we find that the bias of the Schechter model does only slightly grow with increasing resolution while the logarithmic model is more strongly affected.

### Information content

To assess on how much additional information the hierarchical model Eq. (4.1) contains in comparison to a traditional analysis in terms of the tracer power spectra we compare the reconstruction confidence of the power spectrum amplitude chain $\mathcal{C}_A$ to the expected variance of the measured power spectra. The first quantity can simply be determined by averaging the signal-

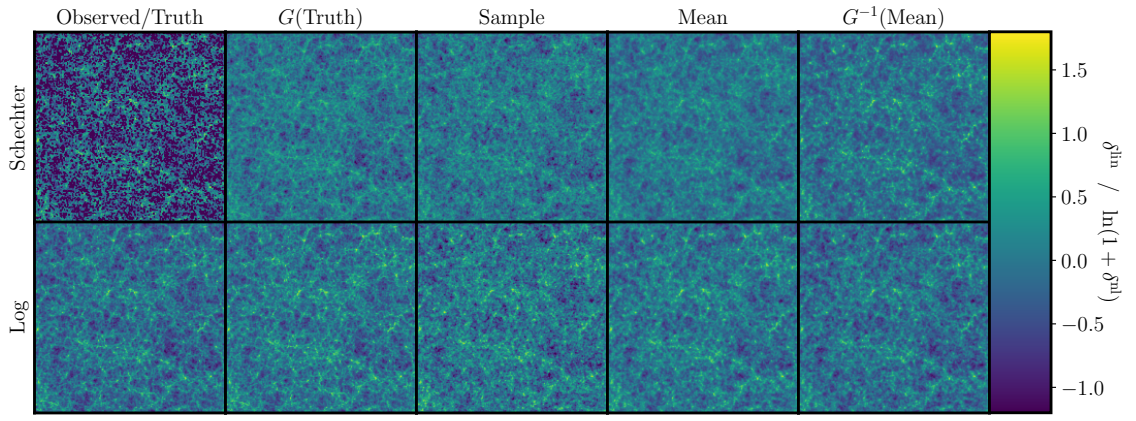


Figure 4.3 Latent field reconstruction on a Quijote slab within the Schechter_BaseRes ensemble. The first column shows the observed tracer counts (top) and the underlying true mass field (bottom). The remaining columns show the linearised mass field (second), a sample from the chain (third), and the mean of all samples for the linear (fourth) and nonlinear (fifth) fields. Panels in the top row were obtained when employing the Schechter transformation Eq. (4.7) while the bottom row was the result of assuming a lognormal model Eq. (4.6). for $G$. For both runs we chose $\bar{n} = 0.002/\mathrm{Mpc}^3$ and matched all the random seeds.

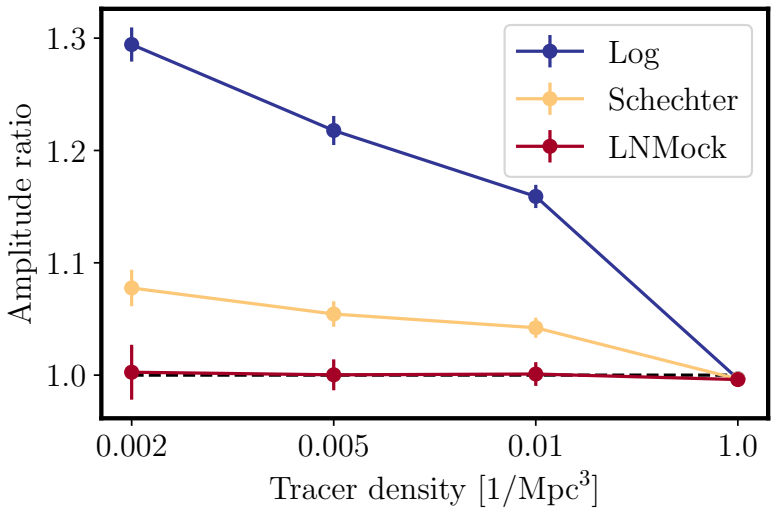


Figure 4.4 Ratio of the reconstructed power spectrum amplitudes as compared to the best fit value $A^*$ across the ensemble for different tracer overdensities.

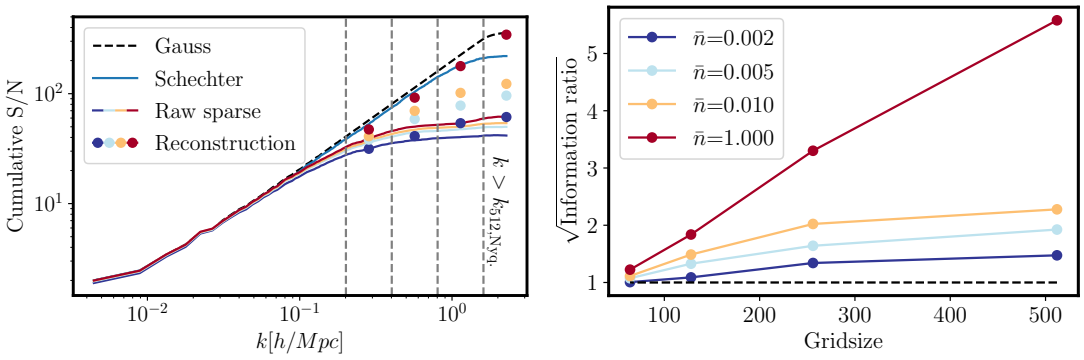


Figure 4.5 *Left hand side*: Cumulative signal-to-noise ratio on the linear power spectrum amplitude $A$ when using the power spectrum (lines) or the hierarchical model Eq. (4.1) with the Schechter transformation (dots). The four different colors indicate the sampling density of the tracers (see legend on the right hand side plot) and the different sets of points correspond to the results when running the reconstruction method on different resolutions, see Table 4.1 for the details. The blue line indicates the cumulative signal-to-noise for the transformed field and the dashed black line gives the ideal scaling that would be obtained for a fully Gaussian field. *Right hand side*: Information gain of the reconstruction method with respect to the traditional one.

to-noise of the $\mathcal{C}_A$ over the ensemble,

$$\left(\frac{S}{N}\right)_{\text{hier.}} \equiv \left\langle \frac{\mathbb{E}(\mathcal{C}_A)}{\sigma(\mathcal{C}_A)} \right\rangle_{\text{chains}} . \tag{4.18}$$

For obtaining the corresponding measure from the standard analysis we follow the procedure put forward in Rimes & Hamilton (2005); Lee & Pen (2008) and estimate the cumulative information content of the projected power spectra,

$$\left(\frac{S}{N}\right)^2 (< k) \equiv \sum_{i,j \in \mathcal{R}_k} r_{ij}^{-1} , \tag{4.19}$$

where $r$ denotes the normalised correlation matrix of the shot-noise subtracted spectra and the sum runs over all index pairs for which $k_{i,j} < k$. To minimize the noise when estimating $r$ and therefore get a stable inverse we obtain this quantity from 2000 Quijote simulations which were run on the standard resolution, i.e. using $512^3$ dark matter particles from which we construct the slabs and tracer counts as described in §4.2.

In Fig. 4.5 we compare the two measures Eqs. (4.18) and (4.19) for the Schechter models of different resolutions. Firstly, we see that that the cumulative signal-to-noise of the traditional analysis starts to diverge from the Gaussian scaling at around $k = 0.1h/\mathrm{Mpc}$ and quickly plateaus thereafter. In contrast, for the hierarchical model the confidence on the linear amplitude has a significant dependence on the tracer density and resolution and does always yield more

information than the power spectrum based method. For a high-sampling limit of $\bar{n} = 1h/\mathrm{Mpc}^3$ Eq. (4.1) recovers the Gaussian mode count up until scales down to $2\mathrm{Mpc}/h$. In the plot on the right we quantify the information ratio of both analysis methods and we see that there is a significant advantage for transverse scales $< 8\mathrm{Mpc}/h$. In particular, the information is quadrupled at translinear scales for an attainable shot-noise level of $\bar{n} = 0.01h/\mathrm{Mpc}^3$.

## 4.5 Conclusions

In this chapter we have explored the information content of a simplistic forward model for projected galaxy fields. It can be summarized as follows:

In §4.2 we wrote down a hierarchical model that is composed of a Poisson likelihood for the discrete galaxy distribution and a Gaussian prior on the linearized density field. We showed that a Schechter like functional form is sufficient to parametrize the gaussianizing point transformation of the nonlinear density field. In §4.3 we introduced the HMC sampling algorithm and wrote down the explicit expression for the potential corresponding to the hierarchical model. We then gave an overview of our implementation choices, namely a non-diagonal mass matrix, a second order Leapfrog integration scheme, an automatic stepsize adaption algorithm, as well as some storage optimizations. In §4.4 we started by laying out our nested burnin strategy and gave an overview of the four hyperparameters that are varied during the ensemble runs. After having successfully validated our implementation on lognormal mock catalogs we presented a summary of our ensemble runs on the Quijote simulation suite. Analyzing the resulting chains we found that the reconstruction bias of the power spectrum amplitude depends on the chosen functional form of the linearization procedure. Comparing the reconstruction confidence of the hierarchical model to the cumulative information content of the raw field we showed that the forward modelling approach allows to extract four times the information as compared to traditional analysis methods.

# Appendix

## 4.A Details of the implementation

### Choice of mass matrix

The mass matrix can be written in a block like structure consisting of the the Hessian associated with the latent field parameters, the Hessian with respect to the power spectrum amplitude, and a mixed one. For this work we make the following choices:

$$M = \left[\begin{array}{c|c} \nabla^2_{\boldsymbol{\delta}} \left(\psi_{\text{Poiss}} + \psi_{\text{Gauss}}\right) & \nabla^2_{\boldsymbol{\delta},\boldsymbol{\Pi}_{P^*}} \psi_{\text{Gauss}} \\ \hline \nabla^2_{\boldsymbol{\delta},\boldsymbol{\Pi}_{P^*}} \psi_{\text{Gauss}} & \nabla^2_{\boldsymbol{\Pi}_{P^*}} \psi_{\text{Gauss}} \end{array}\right] \equiv \left[\begin{array}{c|c} \nabla^2_{\boldsymbol{a}} \psi_{\text{Gauss}} + \boldsymbol{\varepsilon} & -\frac{\partial \boldsymbol{P}/\partial \boldsymbol{\Pi}_{P^*}}{\boldsymbol{P}^{*3/2}} \\ \hline -\frac{\partial \boldsymbol{P}/\partial \boldsymbol{\Pi}_{P^*}}{\boldsymbol{P}^{*3/2}} & \frac{1}{2} \frac{\partial \boldsymbol{P}/\partial \boldsymbol{\Pi}_{P^*}}{\boldsymbol{P}^*} \cdot \frac{\partial \boldsymbol{P}/\partial \boldsymbol{\Pi}_{P^*}}{\boldsymbol{P}^*} \end{array}\right]$$
$$= \left[\begin{array}{c|c} \frac{\mathbb{1}}{\boldsymbol{P}^*} \left(\mathbf{1} + \boldsymbol{P}^* \odot \boldsymbol{\varepsilon}\right) & -\frac{1}{A^* \sqrt{\boldsymbol{P}^*}} \\ \hline -\left(\frac{1}{A^* \sqrt{\boldsymbol{P}^*}}\right)^T & \frac{n_L}{2 A^{*2}} \end{array}\right] , \tag{4.20}$$

where in the first step we dropped the Poisson contribution, introduced a positive definite, constant normalization vector $\boldsymbol{\varepsilon}$ that is needed to keep the mass well defined and substituted the Fourier space representation of Eq. (4.9) evaluated at $a^\dagger a = P^*$. In the second step we specified to the case in which $\boldsymbol{\Pi}^*_P \equiv A^*$. As the latent part has a diagonal structure we can easily compute the inverse with help of the block matrix inversion formula:

$$\boldsymbol{M}^{-1} = \left[\begin{array}{c|c} \left(\sqrt{\boldsymbol{P}^*} \odot \boldsymbol{X}\right) \mathbb{1} + \alpha \boldsymbol{X} \otimes \boldsymbol{X} & \alpha A^* \boldsymbol{X} \\ \hline \alpha A^* \boldsymbol{X}^T & \alpha A^{*2} \end{array}\right] ,$$
$$\text{where} \qquad \boldsymbol{X} \equiv \frac{\sqrt{\boldsymbol{P}^*}}{\mathbf{1} + \boldsymbol{P}^* \odot \boldsymbol{\varepsilon}} \qquad \alpha \equiv \frac{2}{n_L} \frac{1}{1 - 2 \left\langle \frac{1}{1 + \boldsymbol{P}^* \odot \boldsymbol{\varepsilon}} \right\rangle} . \tag{4.21}$$

To circumvent the large storage requirements of this representation we make use the fact that we only need to perform the operation $\boldsymbol{M}^{-1}\boldsymbol{p}$ within the integrator which can be done in linear

time and space complexity as

$$\boldsymbol{M}^{-1}\left[\frac{\boldsymbol{p}_L}{\boldsymbol{p}_{\boldsymbol{\Pi}}}\right] = \left[\frac{\sqrt{\boldsymbol{P}^*}\odot\boldsymbol{X}\odot\boldsymbol{p}_L + \alpha\Sigma\boldsymbol{X} + \alpha A^* p_\Pi \boldsymbol{X}}{\alpha A^*\Sigma + A^{*2}\alpha p_\Pi}\right], \tag{4.22}$$

where we defined $\Sigma \equiv \boldsymbol{X}^T \cdot \boldsymbol{p}_L$. We can now draw a random momentum vector as follows:

1. Compute a lower triangular matrix $\boldsymbol{L}$ s.t. $\boldsymbol{M} = \boldsymbol{L}\boldsymbol{L}^T$. In our case the resulting Cholesky matrix reads
$$L = \left[\begin{array}{c|c} \frac{\mathbb{1}}{\sqrt{\boldsymbol{P}^*}}\sqrt{\mathbf{1}+\boldsymbol{P}^*\odot\boldsymbol{\varepsilon}} & 0 \\ \hline -\left(\frac{\boldsymbol{X}}{A^*\sqrt{\boldsymbol{P}^*}}\right)^T & \frac{1}{\sqrt{\alpha}A^*} \end{array}\right]. \tag{4.23}$$
2. Draw a random unit Gaussian, $\boldsymbol{z} \sim \mathcal{H}(0, \mathbb{1}_{n_L}) \oplus \mathcal{G}(0,1)$, where for the latent field pixels we enforce hermitian symmetry.
3. Transform $z$ to inherit the correct covariance properties: $\boldsymbol{p} = \boldsymbol{L}\boldsymbol{z}$. This can again be done without needing to store $\boldsymbol{L}$ as a whole.

### Dual Averaging

We implemented a dual averaging algorithm using the same formalism as presented in Hoffman & Gelman (2014), see their Algorithm 5 for the specific implementation and initialization values. As our target acceptance rate we chose $\delta = 0.65$. In Fig. 4.6 we show how well this setup did work in the ensemble runs which were all starting the first burnin stage with a the same, very small stepsize. We see that the resulting stepsize is indeed compatible with the target acceptance rate for all the probed ensemble runs. We also see that there is a strong dependence of the inferred stepsize on the galaxy sampling density and the pixel scale.

However, we also found that for this automatically inferred stepsize there is a non-negligible amplitude bias present in the lognormal mocks for a very low galaxy sampling density. When reducing the stepsize by a factor of three[3] this bias disappeared, see Fig. 4.7. While we did not find such a strong change for the Schechter transformation we think that one should be aware of this effect for more complex hierarchical models.

## 4.B Convergence tests

In Fig. 4.8 we show the Gelman Rubin diagnostics Eq. (1.44) for the Schechter_BaseRes run. We see that the amplitude chains did converge for practically all our runs, except for some recon-

[3] We adapt the the number of integration steps accordingly to have a similar trajectory length for each of the runs.

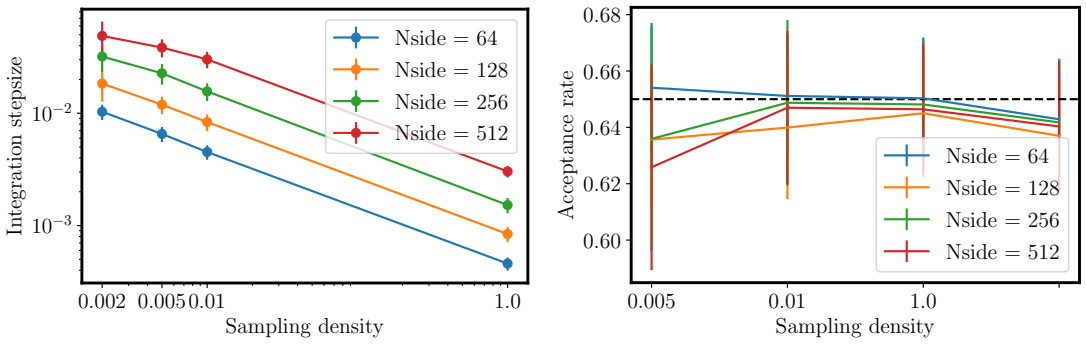


Figure 4.6 *Left hand side*: The final stepsize proposed by the dual averaging algorithm for the Schechter models. *Right hand side*: Mean acceptance rate of the chains during the sampling stage. The black dashed line indicates the target acceptance rate of $\delta \equiv 0.65$ used within the dual averaging algorithm. For both plots we took into account all the chains in the Schechter models and averaged over all the seeds.

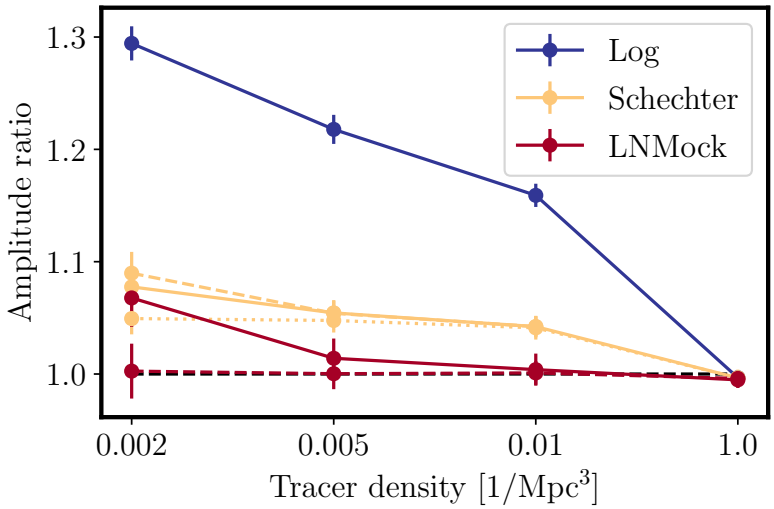


Figure 4.7 Dependence of the power spectrum amplitude ratio on the chosen stepsize for different transformations in the ensemble runs performed on a $256^2$ grid. The solid lines display the results for the unmodified stepsize $\epsilon_{\max}$ while for the dashed and dotted lines we choose $\epsilon_{\max}/3$ and $\epsilon_{\max}/10$.

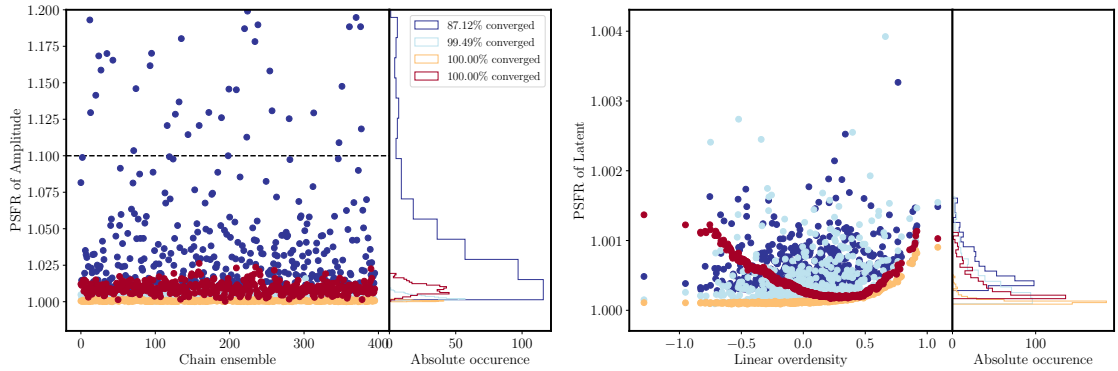


Figure 4.8 Convergence diagnostic of the chains in the Quijote baseline ensemble run. The left hand side shows the results for the power spectrum amplitude for each of the 400 individual runs. The color coding corresponds to the four different galaxy sampling densities. On the right hand side we plot the ensemble mean of the PSFR for 500 randomly chosen pixels.

structions performed with very low galaxy sampling densities. However, as those PSFR values are only slightly above the 1.1 level and we just care about the ensemble mean of the distributions, this should not pose a problem. For the latent pixel parameters we see that the mean PSFR for all the runs is very well consistent with unity. Checking the Gelman Rubin diagnostic on the other ensemble runs we find similar results.

# Summary and Outlook

# 5

In this thesis we have explored two different methods that both aim at extracting cosmological information that is not contained in the power spectrum. Our contributions can be summarized as follows:

In chapter 2 we revisited the originally proposed direct estimator of the second order aperture mass statistics of Schneider et al. (1998) and modified it in two ways. On the one hand, we rewrote the expressions in a fashion that facilitates a fast computational scaling. On the other hand, we proposed a weighting scheme that is based on the shape noise dominated limit of the inverse variance of the second order estimator and that additionally allows to fully exclude apertures that are partially covered by a survey mask. We then compared our modified direct estimator to the traditional analysis method in terms of shear correlation functions on a large suite of ray-tracing mock simulations. We found that the novel weighting scheme does eliminate most of the mask induced bias while still maintaining a high signal-to-noise for an aperture completeness cutoff of $0.7$. We used this observation to apply our direct estimator to the CFHTLenS data and found it to be in excellent agreement with the published results from Kilbinger et al. (2013). Finally, we used the Fisher matrix formalism and a large suite of mock catalogs that were constructed from simulations spanning multiple cosmologies to conclude that the information content of our direct estimator and the correlation function method is comparable.

In chapter 3 we extended the direct estimator put forward in chapter 2 to arbitrary multiscale aperture measures. As those statistics trace the full information probed by the polyspectra they are expected to strongly help in obtaining tighter parameter constraints. After having motivated a linear order algorithm for estimating the statistics we computed its variance for an ensemble of apertures and we showed that the spatial correlation between neighbouring apertures induces a limit for the aperture oversampling rate above which the information saturates. After having

successfully validated our implementation we made use of the SLICS simulation suite and showed that for a KiDS-1000 like survey one can expect to extract information up to fifth order for both, single and multiscale apertures.

An immediate extension of the presented work would be to theoretically assess the impact of a survey mask when adopting the direct estimator approach. This is an ongoing project and we have developed exact theoretical expressions for the expected measurement on an incomplete aperture and have furthermore worked out template based approximation schemes that make the calculations feasible when dealing with ensembles of thousands of apertures. In particular, we have explicitly checked that our method is able recover the second order masked statistics at the per cent level. For an overview of those preliminary results, see Fig. 5.1. We plan on extending these methods to the higher order multiscale apertures.

Up to now we have been concerned with extracting cosmological information from moments of the mass field. While under the assumption of a nearly Gaussian CMB statistical physics would guarantee most of the information to be contained in the lower order moments, we will still never be able to extract the full information of the nonlinear density field using such methods. In chapter 4 we therefore shifted gears and made use of a forward modelling formalism for projected galaxy fields that uses the full observed data instead of summary statistics. To build a simplified model we assumed that structure formation can be described as a point transformation and that galaxies are Poisson tracers of the dark matter field. We used the Quijote simulation suite to determine the one point gaussianizing transformation of the mass field and showed that its inverse differs substantially from a logarithmic model and is parameterizable by a Schechter-like function. By adopting a Hamiltonian Monte Carlo sampling scheme we were able to efficiently extract the information of the power spectrum amplitude of the linearized mass field from the high dimensional posterior. Comparing this information to the expected one when adopting a standard power spectrum analysis on the raw field we found that the forward model is able to produce four times tighter constraints.

There are multiple ways to extend our analysis. On the modelling side, we could use a more advanced structure formation model beyond a point transformation. As those models are usually formulated in terms of three dimensional fields one could try to find a way to phrase them in terms of projected fields, such that the sampling will remain fast. As long as the adopted model is deterministic, we expect the information content to be very similar to the one found for our simple model. Another immediate extension would be to tailor the formulation towards weak lensing analyses, in which, on a high level, the shot noise is replaced by the shape noise. Given our findings for projected fields we would expect the information gain of a hierarchical model

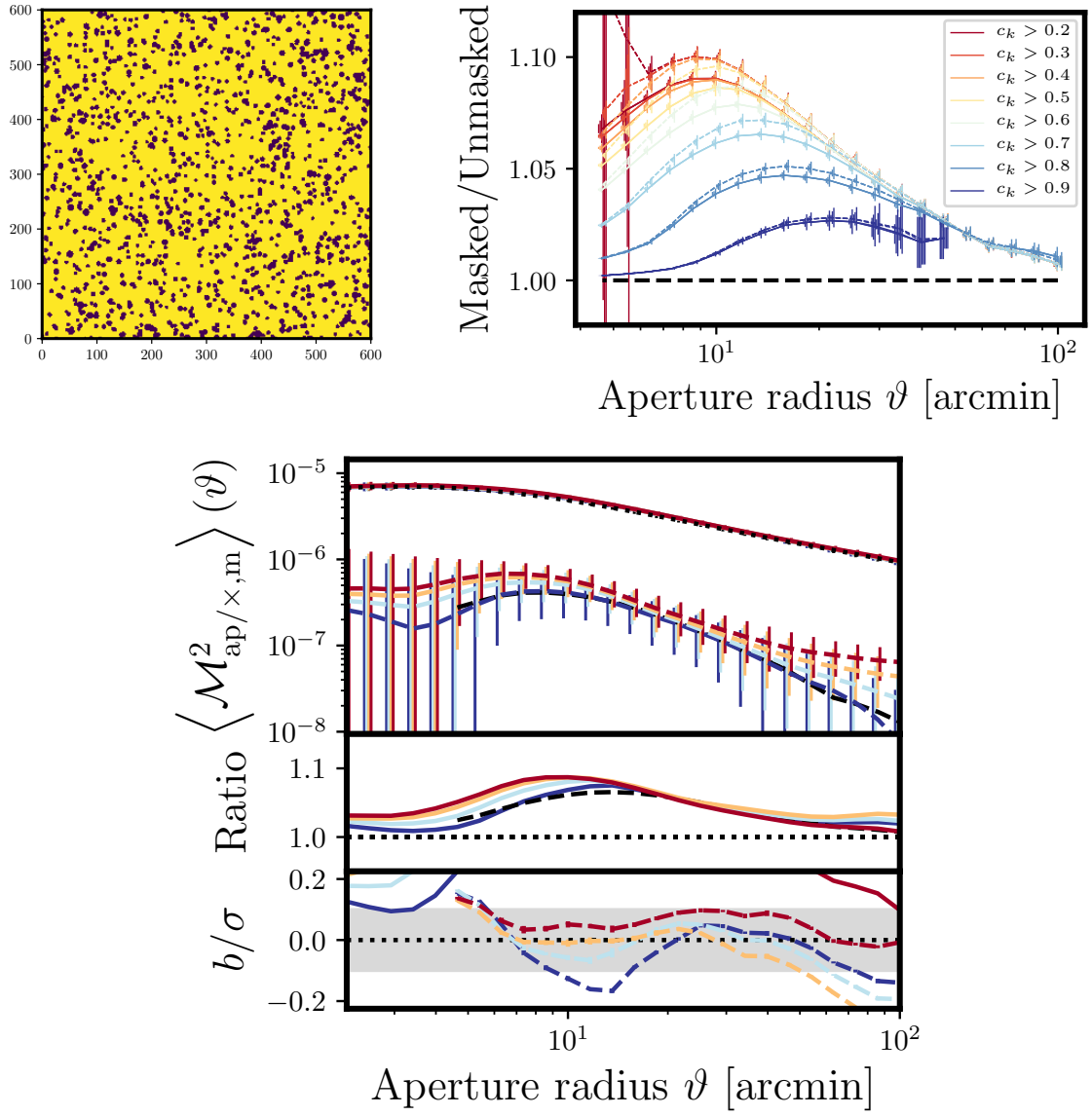


Figure 5.1 *Upper left hand side*: The angular mask adopted for this example. *Upper right hand side*: The theoretically predicted impact of the mask on the second order E-mode aperture statistics for two different analytic approximation schemes. The different colors indicate different aperture completeness cutoffs $c_k$. *Lower panel*: Comparison of the theoretical calculations to measurements on the SLICS ensemble for cutoff scales of $c_k \in \{0.7, 0.6, 0.5, 0.4\}$. In the upper panel the solid and dashed lines show the measured E and B-mode in the presence of the survey mask. We plot our theoretical prediction for this mode with $c_k = 0.7$ as the black dashed line. In the middle panel we show the ratio between the masked and unmasked E-mode statistics and again compare against the theoretical expectation with $c_k = 0.7$ displayed via the black dashed line. In the lower panel we show the discrepancy of the analytical predictions and the measurements when normalizing by the expected estimation uncertainty for a KiDS-1000 setting.

to be similar to the results presented in this thesis. On the sampling side there are still many different avenues that can be explored. For example, one could investigate the performance increment when using more advanced schemes, such as NUTS or RMHMC.

The work presented in this thesis has contributed to one of the many challenges that need to be tackled in order to optimally use the information collected by ongoing and future surveys. This will ultimately lead to a better understanding of our Universe. In recent years it has become apparent that many of those challenges cannot be treated separately, but instead require a collaborative effort between multiple previously not strongly connected fields of research. I am convinced that this will lead to many new interesting results and that the field of observational cosmology will remain as exciting as it is now!

# Bibliography

Aihara, H., N. Arimoto, R. Armstrong, S. Arnouts, N. A. Bahcall, S. Bickerton, J. Bosch, K. Bundy, P. L. Capak, J. H. H. Chan et al., 2018. The Hyper Suprime-Cam SSP Survey: Overview and survey design. *Publications of the ASJ* **70**:S4. (Cited on pages 31 and 79.)

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang et al., 2017. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Monthly Notices of the Royal Astronomical Society* **470**:2617–2652. (Cited on page 2.)

Albrecht, A. and P. J. Steinhardt, 1982. Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking. *Physical Review Letters* **48**(17):1220–1223. (Cited on page 6.)

Asgari, M., C.-A. Lin, B. Joachimi, B. Giblin, C. Heymans, H. Hildebrandt, A. Kannawadi, B. Stölzner, T. Tröster, J. L. van den Busch et al., 2021. Kids-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. Astronomy Astrophysics **645**:A104. URL http://dx.doi.org/10.1051/0004-6361/202039070. (Cited on page 79.)

Bacon, D. J., A. R. Refregier and R. S. Ellis, 2000. Detection of weak gravitational lensing by large-scale structure. *Monthly Notices of the Royal Astronomical Society* **318**(2):625–640. (Cited on pages 31 and 78.)

Bardeen, J. M., 1980. Gauge-invariant cosmological perturbations. *Physical Review D* **22**(8):1882–1905. (Cited on page 7.)

Bartelmann, M. and P. Schneider, 2001. Weak gravitational lensing. Phys. Rep. **340**:291–472. (Cited on pages 1, 27, 34, 36, 37, 81, and 82.)

Barthelemy, A., S. Codis and F. Bernardeau, 2020. Probability distribution function of the aperture mass field with large deviation theory. (Cited on page 80.)

Baumann, D., 2009. TASI Lectures on Inflation. arXiv e-prints arXiv:0907.5424. (Cited on page 10.)

Benitez, N., 2000. Bayesian Photometric Redshift Estimation. *The Astrophysical Journal* **536**(2):571–583. (Cited on page 19.)

Bernardeau, F., S. Colombi, E. Gaztañaga and R. Scoccimarro, 2002. Large-scale structure of the Universe and cosmological perturbation theory. Phys. Rep. **367**:1–3. (Cited on pages 31, 79, and 96.)

Bernardeau, F., M. Crocce and R. Scoccimarro, 2008. Multipoint propagators in cosmological gravitational instability. *Physical Review D* **78**(10):103521–+. (Cited on page 10.)

Bernardeau, F. and P. Valageas, 2000. Construction of the one-point PDF of the local aperture mass in weak lensing maps. *Astronomy & Astrophysics* **364**:1–16. (Cited on page 80.)

Bertschinger, E., 1995. COSMICS: Cosmological Initial Conditions and Microwave Anisotropy Codes. arXiv e-prints astro-ph/9506070. (Cited on page 8.)

Betancourt, M., 2017. A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv e-prints arXiv:1701.02434. (Cited on page 22.)

Betancourt, M. J., S. Byrne, S. Livingstone and M. Girolami, 2014. The Geometric Foundations of Hamiltonian Monte Carlo. arXiv e-prints arXiv:1410.5110. (Cited on pages 22 and 23.)

Bird, S., M. Viel and M. G. Haehnelt, 2012. Massive neutrinos and the non-linear matter power spectrum. *Monthly Notices of the Royal Astronomical Society* **420**:2551–2561. (Cited on page 13.)

Blandford, R. D., A. B. Saust, T. G. Brainerd and J. V. Villumsen, 1991. The distortion of distant galaxy images by large-scale structure. *Monthly Notices of the Royal Astronomical Society* **251**:600. (Cited on pages 30, 34, and 78.)

Blanes, S. and F. Casas, 2016. A Concise Introduction to Geometric Numerical Integration (1st ed.). Chapman and Hall CRC. (Cited on page 23.)

Blas, D., J. Lesgourgues and T. Tram, 2011. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *Journal of Cosmology and Astro-Particle Physics* **2011**(7):034. (Cited on pages 8 and 10.)

Blazek, J. A., N. MacCrann, M. A. Troxel and X. Fang, 2019. Beyond linear galaxy alignments. *Physical Review D* **100**(10):103506. (Cited on page 19.)

Blei, D. M., A. Kucukelbir and J. D. McAuliffe, 2017. Variational inference: A review for statisticians. Journal of the American Statistical Association **112**(518):859–877. URL https://doi.org/10.1080/01621459.2017.1285773. (Cited on page 22.)

Bolzonella, M., J. M. Miralles and R. Pelló, 2000. Photometric redshifts based on standard SED fitting procedures. *Astronomy & Astrophysics* **363**:476–492. (Cited on page 19.)

Bond, J. R., G. Efstathiou and J. Silk, 1980. Massive neutrinos and the large-scale structure of the universe. Physical Review Letters **45**:1980–1984. (Cited on page 9.)

Byun, J., A. Eggemeier, D. Regan, D. Seery and R. E. Smith, 2017. Towards optimal cosmological parameter recovery from compressed bispectrum statistics. *Monthly Notices of the Royal Astronomical Society* **471**(2):1581–1618. (Cited on pages 31 and 79.)

Carrasco, J. J. M., S. Foreman, D. Green and L. Senatore, 2014. The Effective Field Theory of Large Scale Structures at two loops. *Journal of Cosmology and Astro-Particle Physics* **7**:057. (Cited on page 10.)

Carrasco Kind, M. and R. J. Brunner, 2014. SOMz: photometric redshift PDFs with self-organizing maps and random atlas. *Monthly Notices of the Royal Astronomical Society* **438**(4):3409–3421. (Cited on page 19.)

Catelan, P. and C. Porciani, 2001. Correlations of cosmic tidal fields. *Monthly Notices of the Royal Astronomical Society* **323**(3):713–717. (Cited on page 19.)

Chisari, N. E., D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu, A. Villarreal, S. Singh, T. McClintock, J. Ellison et al., 2019. Core Cosmology Library: Precision Cosmological Predictions for LSST. *The Astrophysical Journal Suppl.* **242**(1):2. (Cited on page 97.)

Coles, P. and B. Jones, 1991. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248**:1–13. (Cited on page 10.)

Collister, A. A. and O. Lahav, 2004. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. **116**(818):345–351. (Cited on page 19.)

Comtet, L., 1974. Advanced Combinatorics: The Art of Finite and Infinite Expansions. Springer Netherlands. (Cited on page 91.)

Cooray, A. and R. Sheth, 2002. Halo models of large scale structure. Phys. Rep. **372**:1–129. (Cited on page 13.)

Cox, R. T., 1946. Probability, frequency, and reasonable expectation. American Journal of Physics **14**(2):1–13. (Cited on page 20.)

Creutz, M. and A. Gocksch, 1989. Higher-order hybrid monte carlo algorithms. Phys. Rev. Lett. **63**:9–12. URL https://link.aps.org/doi/10.1103/PhysRevLett.63.9. (Cited on pages 25 and 135.)

Crittenden, R. G., P. Natarajan, U.-L. Pen and T. Theuns, 2001. Spin-induced Galaxy Alignments and Their Implications for Weak-Lensing Measurements. *The Astrophysical Journal* **559**(2):552–571. (Cited on pages 32, 46, 79, and 100.)

Crocce, M., S. Pueblas and R. Scoccimarro, 2006. Transients from initial conditions in cosmological simulations. *Monthly Notices of the Royal Astronomical Society* **373**:369–381. (Cited on pages 10 and 49.)

Dodelson, S., 2003. Modern cosmology. Academic Press. (Cited on pages 1, 34, and 81.)

Dodelson, S., 2017. Gravitational Lensing. Cambridge University Press. (Cited on pages 34 and 81.)

Dolag, K., S. Borgani, S. Schindler, A. Diaferio and A. M. Bykov, 2008. Simulation Techniques for Cosmological Simulations. **134**(1-4):229–268. (Cited on page 11.)

Duane, S., A. Kennedy, B. J. Pendleton and D. Roweth, 1987. Hybrid monte carlo. Physics Letters B **195**(2):216–222. URL https://www.sciencedirect.com/science/article/pii/037026938791197X. (Cited on pages 22, 28, and 134.)

Durrer, R., 2008. The Cosmic Microwave Background. (Cited on page 7.)

Efstathiou, G. and S. Gratton, 2019. A Detailed Description of the CamSpec Likelihood Pipeline and a Reanalysis of the Planck High Frequency Maps. arXiv e-prints arXiv:1910.00483. (Cited on page 2.)

Einstein, A., 1915. Die Feldgleichungen der Gravitation. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin pages 844–847. (Cited on page 2.)

Fabbian, G., M. Calabrese and C. Carbone, 2018. CMB weak-lensing beyond the Born approximation: a numerical approach. *Journal of Cosmology and Astro-Particle Physics* **2018**(2):050. (Cited on page 79.)

Fang, X., E. Krause, T. Eifler and N. MacCrann, 2020. Beyond Limber: efficient computation of angular power spectra for galaxy clustering and weak lensing. *Journal of Cosmology and Astro-Particle Physics* **2020**(5):010. (Cited on page 18.)

Friedman, A., 1922. Über die Krümmung des Raumes. Zeitschrift für Physik **10**:377–386. (Cited on page 2.)

Friedrich, O., S. Seitz, T. F. Eifler and D. Gruen, 2016. Performance of internal covariance estimators for cosmic shear correlation functions. *Monthly Notices of the Royal Astronomical Society* **456**(3):2662–2680. (Cited on pages 33 and 80.)

Fry, J. N., 1984. The Galaxy correlation hierarchy in perturbation theory. *The Astrophysical Journal* **279**:499–510. (Cited on page 10.)

Fu, L., M. Kilbinger, T. Erben, C. Heymans, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, E. Semboloni et al., 2014. CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *Monthly Notices of the Royal Astronomical Society* **441**(3):2725–2743. (Cited on pages 32, 33, 79, 80, and 92.)

Gelman, A. and D. B. Rubin, 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science **7**(4):457 – 472. URL https://doi.org/10.1214/ss/1177011136. (Cited on pages 22 and 137.)

Geman, S. and D. Geman, 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-6**(6):721–741. (Cited on page 22.)

Girolami, M. and B. Calderhead, 2011. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73**(2):123–214. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x. (Cited on page 25.)

Goodman, J. and J. Weare, 2010. Ensemble samplers with affine invariance. Communications in Applied Mathematics and Computational Science **5**(1):65–80. (Cited on page 22.)

Goroff, M. H., B. Grinstein, S.-J. Rey and M. B. Wise, 1986. Coupling of modes of cosmological mass density fluctuations. *The Astrophysical Journal* **311**:6–14. (Cited on page 10.)

Guth, A. H., 1981. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D* **23**(2):347–356. (Cited on page 6.)

Harnois-Déraps, J., A. Amon, A. Choi, V. Demchenko, C. Heymans, A. Kannawadi, R. Nakajima, E. Sirks, L. van Waerbeke, Y.-C. Cai et al., 2018. Cosmological simulations for combined-probe analyses: covariance and neighbour-exclusion bias. *Monthly Notices of the Royal Astronomical Society* **481**(1):1337–1367. (Cited on pages 96 and 97.)

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Ralf Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith et al., 2020. Array programming with NumPy. Nature **585**(7825):357–362. URL https://doi.org/10.1038/s41586-020-2649-2. (Cited on page 103.)

Harrison, E. R., 1970. Fluctuations at the Threshold of Classical Cosmology. *Physical Review D* **1**(10):2726–2730. (Cited on page 13.)

Hartlap, J., P. Simon and P. Schneider, 2007. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *Astronomy & Astrophysics* **464**:399–404. (Cited on page 62.)

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1):97–109. URL https://doi.org/10.1093/biomet/57.1.97. (Cited on page 21.)

Heavens, A., 2009. Statistical techniques in cosmology. ArXiv e-prints . (Cited on page 1.)

Hernández-Sánchez, M., F.-S. Kitaura, M. Ata and C. Dalla Vecchia, 2021. Higher Order Hamiltonian Monte Carlo Sampling for Cosmological Large-Scale Structure Analysis. *Monthly Notices of the Royal Astronomical Society* **502**(3):3976–3992. URL https://doi.org/10.1093/mnras/stab123. (Cited on page 135.)

Hetterscheidt, M., T. Erben, P. Schneider, R. Maoli, L. Van Waerbeke and Y. Mellier, 2005. Searching for galaxy clusters using the aperture mass statistics in 50 vlt fields. A&A **442**(1):43–61. URL https://doi.org/10.1051/0004-6361:20053339. (Cited on page 50.)

Heydenreich, S., B. Brück and J. Harnois-Déraps, 2020. Persistent homology in cosmic shear: constraining parameters with topological data analysis. (Cited on page 80.)

Heymans, C., L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, P. Simon, C. Bonnett et al., 2012. CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society* **427**(1):146–166. (Cited on page 48.)

Hikage, C., M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada, F. Köhlinger, H. Miyatake, A. J. Nishizawa, H. Aihara et al., 2019. Cosmology from cosmic shear power spectra with

Subaru Hyper Suprime-Cam first-year data. *Publications of the ASJ* **71**(2):43. (Cited on pages 31 and 79.)

Hilbert, S., A. Barreira, G. Fabbian, P. Fosalba, C. Giocoli, S. Bose, M. Calabrese, C. Carbone, C. T. Davies, B. Li et al., 2020. The accuracy of weak lensing simulations. Monthly Notices of the Royal Astronomical Society **493**(1):305–319. URL https://doi.org/10.1093/mnras/staa281. (Cited on page 54.)

Hilbert, S., J. Hartlap, S. D. M. White and P. Schneider, 2009. Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *Astronomy & Astrophysics* **499**:31–43. (Cited on pages 31, 50, and 79.)

Hilbert, S., L. Marian, R. E. Smith and V. Desjacques, 2012. Measuring primordial non-Gaussianity with weak lensing surveys. *Monthly Notices of the Royal Astronomical Society* **426**(4):2870–2888. (Cited on pages 31 and 79.)

Hilbert, S., R. B. Metcalf and S. D. M. White, 2007. Imaging the cosmic matter distribution using gravitational lensing of pre-galactic HI. *Monthly Notices of the Royal Astronomical Society* **382**(4):1494–1502. (Cited on page 50.)

Hildebrandt, H., M. Viola, C. Heymans, S. Joudaki, K. Kuijken, C. Blake, T. Erben, B. Joachimi, D. Klaes, L. Miller et al., 2017. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society* **465**(2):1454–1498. (Cited on pages 31, 79, and 96.)

Hirata, C. M., R. Mandelbaum, M. Ishak, U. Seljak, R. Nichol, K. A. Pimbblet, N. P. Ross and D. Wake, 2007. Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: luminosity and redshift scalings and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society* **381**(3):1197–1218. (Cited on page 19.)

Hoffman, M. D. and A. Gelman, 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. arXiv e-prints arXiv:1111.4246. (Cited on page 25.)

Hoffman, M. D. and A. Gelman, 2014. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research **15**(47):1593–1623. URL http://jmlr.org/papers/v15/hoffman14a.html. (Cited on pages 135 and 145.)

Hogg, D. W., 1999. Distance measures in cosmology. arXiv e-prints astro-ph/9905116. (Cited on page 5.)

Huang, H.-J., T. Eifler, R. Mandelbaum and S. Dodelson, 2019. Modelling baryonic physics in future weak lensing surveys. *Monthly Notices of the Royal Astronomical Society* **488**(2):1652–1678. (Cited on page 14.)

Hunter, J. D., 2007. Matplotlib: A 2d graphics environment. Computing in Science & Engineering **9**(3):90–95. (Cited on pages 67 and 103.)

Ilbert, O., S. Arnouts, H. J. McCracken, M. Bolzonella, E. Bertin, O. Le Fèvre, Y. Mellier, G. Zamorani, R. Pellò, A. Iovino et al., 2006. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *Astronomy & Astrophysics* **457**(3):841–856. (Cited on page 19.)

Isserlis, L., 1918. On a formula for the product-moment coefficient of any order for a normal frequency distribution in any number of variables. Biometrika **12**(1-2):134–139. URL https://doi.org/10.1093/biomet/12.1-2.134. (Cited on page 13.)

Jain, B. and E. Bertschinger, 1994. Second-order power spectrum and nonlinear evolution at high redshift. *The Astrophysical Journal* **431**:495–505. (Cited on page 10.)

Jain, B. and U. Seljak, 1997. Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes. *The Astrophysical Journal* **484**(2):560–573. (Cited on pages 34 and 78.)

Jarvis, M., G. Bernstein and B. Jain, 2004. The skewness of the aperture mass statistic. *Monthly Notices of the Royal Astronomical Society* **352**(1):338–352. (Cited on pages 32, 53, 54, 59, and 80.)

Jarvis, M., G. M. Bernstein, P. Fischer, D. Smith, B. Jain, J. A. Tyson and D. Wittman, 2003. Weak-Lensing Results from the 75 Square Degree Cerro Tololo Inter-American Observatory Survey. **125**:1014–1032. (Cited on pages 80 and 81.)

Jasche, J. and F. S. Kitaura, 2010. Fast Hamiltonian sampling for large-scale structure inference. *Monthly Notices of the Royal Astronomical Society* **407**(1):29–42. (Cited on page 130.)

Jasche, J. and B. D. Wandelt, 2013a. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *Monthly Notices of the Royal Astronomical Society* **432**(2):894–913. (Cited on page 130.)

Jasche, J. and B. D. Wandelt, 2013b. Methods for Bayesian Power Spectrum Inference with Galaxy Surveys. *The Astrophysical Journal* **779**(1):15. (Cited on page 130.)

Jaynes, E. T., 2003. Probability theory: The logic of science. Cambridge University Press, Cambridge. (Cited on page 20.)

Jeffrey, N. and F. B. Abdalla, 2019. Parameter inference and model comparison using theoretical predictions from noisy simulations. *Monthly Notices of the Royal Astronomical Society* **490**(4):5749–5756. (Cited on page 63.)

Joachimi, B., A. N. Taylor and A. Kiessling, 2011. Cosmological information in Gaussianized weak lensing signals. *Monthly Notices of the Royal Astronomical Society* **418**(1):145–169. (Cited on page 130.)

Juszkiewicz, R., 1981. On the evolution of cosmological adiabatic perturbations in the weakly non-linear regime. *Monthly Notices of the Royal Astronomical Society* **197**:931–940. (Cited on page 10.)

Kacprzak, T., D. Kirk, O. Friedrich, A. Amara, A. Refregier, L. Marian, J. P. Dietrich, E. Suchyta, J. Aleksić, D. Bacon et al., 2016. Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data. *Monthly Notices of the Royal Astronomical Society* **463**(4):3653–3673. (Cited on pages 32 and 79.)

Kaiser, N., 1992. Weak gravitational lensing of distant galaxies. *The Astrophysical Journal* **388**:272–286. (Cited on page 18.)

Kaiser, N., 1995. Nonlinear Cluster Lens Reconstruction. *The Astrophysical Journal Lett.* **439**:L1. (Cited on pages 36 and 80.)

Kaiser, N., 1998. Weak Lensing and Cosmology. *The Astrophysical Journal* **498**(1):26–42. (Cited on pages 30, 34, and 78.)

Kaiser, N. and G. Squires, 1993. Mapping the Dark Matter with Weak Gravitational Lensing. *The Astrophysical Journal* **404**:441. (Cited on page 92.)

Kaiser, N., G. Wilson and G. A. Luppino, 2000. Large-Scale Cosmic Shear Measurements. arXiv e-prints astro-ph/0003338. (Cited on pages 31 and 78.)

Kayo, I., M. Takada and B. Jain, 2013. Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix. *Monthly Notices of the Royal Astronomical Society* **429**(1):344–371. (Cited on pages 31 and 79.)

Kazanas, D., 1980. Dynamics of the universe and spontaneous symmetry breaking. *The Astrophysical Journal Lett.* **241**:L59–L63. (Cited on page 6.)

Kilbinger, M., 2015. Cosmology with cosmic shear observations: a review. Reports on Progress in Physics **78**(8):086901. URL https://doi.org/10.1088/0034-4885/78/8/086901. (Cited on pages 1, 15, and 81.)

Kilbinger, M., C. Bonnett and J. Coupon, 2014. athena: Tree code for second-order correlation functions ascl:1402.026. (Cited on page 32.)

Kilbinger, M., L. Fu, C. Heymans, F. Simpson, J. Benjamin, T. Erben, J. Harnois-Déraps, H. Hoekstra, H. Hildebrand t, T. D. Kitching et al., 2013. CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society* **430**(3):2200–2220. (Cited on pages 34, 58, 59, 61, 66, and 148.)

Kilbinger, M., C. Heymans, M. Asgari, S. Joudaki, P. Schneider, P. Simon, L. Van Waerbeke, J. Harnois-Déraps, H. Hildebrandt, F. Köhlinger et al., 2017. Precision calculations of the cosmic shear power spectrum projection. *Monthly Notices of the Royal Astronomical Society* **472**(2):2126–2141. (Cited on pages 18 and 28.)

Kilbinger, M. and P. Schneider, 2005. Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear. *Astronomy & Astrophysics* **442**(1):69–83. (Cited on pages 31, 32, 33, 79, and 80.)

Kilbinger, M., P. Schneider and T. Eifler, 2006. E- and B-mode mixing from incomplete knowledge of the shear correlation. *Astronomy & Astrophysics* **457**(1):15–19. (Cited on pages 40 and 55.)

Kitaura, F. S. and T. A. Enßlin, 2008. Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization. *Monthly Notices of the Royal Astronomical Society* **389**(2):497–544. (Cited on page 130.)

Knuth, D. E., 2005. The Art of Computer Programming, Volume 4, Fascicle 3: Generating All Combinations and Partitions. Addison-Wesley Professional. (Cited on page 91.)

Komatsu, E., J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page et al., 2009. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *The Astrophysical Journal Suppl.* **180**:330–376. (Cited on page 49.)

Krause, E. and C. M. Hirata, 2010. Weak lensing power spectra for precision cosmology. Multiple-deflection, reduced shear, and lensing bias corrections. *Astronomy & Astrophysics* **523**:A28. (Cited on page 18.)

Laureijs, R., J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet et al., 2011. Euclid Definition Study Report. arXiv e-prints arXiv:1110.3193. (Cited on pages 31 and 79.)

Lavaux, G. and J. Jasche, 2016. Unmasking the masked Universe: the 2M++ catalogue through Bayesian eyes. *Monthly Notices of the Royal Astronomical Society* **455**(3):3169–3179. (Cited on page 130.)

Layzer, D., 1956. A new model for the distribution of galaxies in space. **61**:383. (Cited on page 131.)

Leclercq, F., J. Jasche, G. Lavaux, B. Wandelt and W. Percival, 2017. The phase-space structure of nearby dark matter as constrained by the SDSS. *Journal of Cosmology and Astro-Particle Physics* **2017**(6):049. (Cited on page 130.)

Lee, J. and U. Pen, 2008. Information Content in the Galaxy Angular Power Spectrum from the Sloan Digital Sky Survey and Its Implication on Weak-Lensing Analysis. *The Astrophysical Journal Lett.* **686**:L1–L4. (Cited on page 142.)

Lemaître, A. G., 1931. A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulæ. *Monthly Notices of the Royal Astronomical Society* **91**(5):483–490. URL https://doi.org/10.1093/mnras/91.5.483. (Cited on page 2.)

Lemos, P., A. Challinor and G. Efstathiou, 2017. The effect of Limber and flat-sky approximations on galaxy weak lensing. JCAP **05**:014. (Cited on page 28.)

Lesgourgues, J. and S. Pastor, 2006. Massive neutrinos and cosmology. Phys. Rep. **429**:307–379. (Cited on pages 1 and 10.)

Lewis, A., A. Challinor and A. Lasenby, 2000. Efficient computation of CMB anisotropies in closed FRW models. Astrophys. J. **538**:473–476. (Cited on page 8.)

Lifshitz, E. M., 1946. On the gravitational stability of the expanding universe. Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki **16**:587–602. (Cited on page 7.)

Lima, M., C. E. Cunha, H. Oyaizu, J. Frieman, H. Lin and E. S. Sheldon, 2008. Estimating the redshift distribution of photometric galaxy samples. *Monthly Notices of the Royal Astronomical Society* **390**(1):118–130. (Cited on page 19.)

Limber, D. N., 1953. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *The Astrophysical Journal* **117**:134. (Cited on page 18.)

Linde, A. D., 1982. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. Physics Letters B **108**(6):389–393. (Cited on page 6.)

LSST, 2009. LSST Science Book, Version 2.0. ArXiv e-prints page arXiv:0912.0201. (Cited on pages 31 and 79.)

Ma, C.-P. and E. Bertschinger, 1995. Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. *The Astrophysical Journal* **455**:7–+. (Cited on page 8.)

Malz, A. I., 2021. How not to obtain the redshift distribution from probabilistic redshift estimates: Under what conditions is it not inappropriate to estimate the redshift distribution N (z ) by stacking photo-z PDFs? *Physical Review D* **103**(8):083502. (Cited on page 19.)

Mandelbaum, R., 2018. Weak lensing for precision cosmology. Annual Review of Astronomy and Astrophysics **56**(1):393–433. URL https://doi.org/10.1146/annurev-astro-081817-051928. (Cited on pages 1, 19, and 81.)

Marian, L., R. E. Smith, S. Hilbert and P. Schneider, 2012. Optimized detection of shear peaks in weak lensing maps. *Monthly Notices of the Royal Astronomical Society* **423**(2):1711–1725. URL https://doi.org/10.1111/j.1365-2966.2012.20992.x. (Cited on page 80.)

Marian, L., R. E. Smith, S. Hilbert and P. Schneider, 2013. The cosmological information of shear peaks: beyond the abundance. *Monthly Notices of the Royal Astronomical Society* **432**(2):1338–1350. (Cited on pages 31, 32, and 79.)

Martinet, N., J. Harnois-Déraps, E. Jullo and P. Schneider, 2021. Probing dark energy with tomographic weak-lensing aperture mass statistics. (Cited on page 80.)

Massey, R., H. Hoekstra, T. Kitching, J. Rhodes, M. Cropper, J. Amiaux, D. Harvey, Y. Mellier, M. Meneghetti, L. Miller et al., 2013. Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation). *Monthly Notices of the Royal Astronomical Society* **429**(1):661–678. (Cited on pages 31 and 79.)

Mead, A. J., J. A. Peacock, C. Heymans, S. Joudaki and A. F. Heavens, 2015. An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. *Monthly Notices of the Royal Astronomical Society* **454**:1958–1975. (Cited on page 13.)

Menard, B., R. Scranton, S. Schmidt, C. Morrison, D. Jeong, T. Budavari and M. Rahman, 2013. Clustering-based redshift estimation: method and application to data. arXiv e-prints arXiv:1303.4722. (Cited on page 19.)

Meszaros, P., 1974. The behaviour of point masses in an expanding cosmological substratum. *Astronomy & Astrophysics* **37**(2):225–228. (Cited on page 9.)

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953. Equation of State Calculations by Fast Computing Machines. **21**(6):1087–1092. (Cited on page 21.)

Miller, L., C. Heymans, T. D. Kitching, L. van Waerbeke, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, B. T. P. Rowe, J. Coupon et al., 2013. Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society* **429**(4):2858–2880. (Cited on pages 56 and 59.)

Miralda-Escude, J., 1991. The correlation function of galaxy ellipticities produced by gravitational lensing. *The Astrophysical Journal* **380**:1–8. (Cited on pages 31 and 79.)

Mukhanov, V. F. and G. V. Chibisov, 1981. Quantum fluctuations and a nonsingular universe. Soviet Jour. of Exper. and Theor. Phys. Lett. **33**:532–+. (Cited on page 6.)

Mukhanov, V. F., H. A. Feldman and R. H. Brandenberger, 1992. Theory of cosmological perturbations. Phys. Rep. **215**(5-6):203–333. (Cited on page 7.)

Munshi, D. and P. Coles, 2003. Error estimates for measurements of cosmic shear. Monthly Notices of the Royal Astronomical Society **338**(4):846–856. URL https://doi.org/10.1046/j.1365-8711.2003.06136.x. (Cited on pages 84, 85, 88, and 117.)

Munshi, D. and P. Valageas, 2005. On cross-correlating weak lensing surveys. Monthly Notices of the Royal Astronomical Society **356**(2):439–455. URL https://doi.org/10.1111/j.1365-2966.2004.08462.x. (Cited on page 84.)

Munshi, D., P. Valageas and A. J. Barber, 2004. Weak lensing shear and aperture mass from linear to non-linear scales. Monthly Notices of the Royal Astronomical Society **350**(1):77–94. URL https://doi.org/10.1111/j.1365-2966.2004.07553.x. (Cited on page 80.)

Neal, R. M., 2003. Slice sampling. The Annals of Statistics **31**(3):705 – 767. URL https://doi.org/10.1214/aos/1056562461. (Cited on page 22.)

Neal, R. M., 2012. MCMC using Hamiltonian dynamics. arXiv e-prints arXiv:1206.1901. (Cited on pages 22 and 134.)

Nesterov, Y., 2009. Primal-dual subgradient methods for convex problems. Mathematical Programming **120**(1):221—259. URL https://doi.org/10.1007/s10107-007-0149-x. (Cited on page 135.)

Newman, J. A., 2008. Calibrating Redshift Distributions beyond Spectroscopic Limits with Cross-Correlations. *The Astrophysical Journal* **684**(1):88–101. (Cited on page 19.)

Neyman, J. and E. L. Scott, 1952. A Theory of the Spatial Distribution of Galaxies. *The Astrophysical Journal* **116**:144. (Cited on page 13.)

Neyrinck, M. C., I. Szapudi and A. S. Szalay, 2009. Rejuvinating power spectra ii: Restoring information with a logarithmic density mapping. The Astrophysical Journal **698**(2):L90–L93. URL https://doi.org/10.1088/0004-637x/698/2/l90. (Cited on page 130.)

Neyrinck, M. C., I. Szapudi and A. S. Szalay, 2011. Rejuvinating power spectra ii: The gaussianized galaxy density field. The Astrophysical Journal **731**(2):116. URL https://doi.org/10.1088/0004-637x/731/2/116. (Cited on page 130.)

Oliphant, T. E., 2006. A guide to NumPy, volume 1. Trelgol Publishing USA. (Cited on page 67.)

Peebles, P. J. E., 1980. The large-scale structure of the universe. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p. (Cited on page 131.)

Peebles, P. J. E. and J. T. Yu, 1970. Primeval Adiabatic Perturbation in an Expanding Universe. *The Astrophysical Journal* **162**:815. (Cited on page 9.)

Penrose, R., 1989. Difficulties with inflationary cosmology. Annals of the New York Academy of Sciences **571**. (Cited on page 6.)

Piattella, O., 2018. Lecture Notes in Cosmology. Springer International Publishing. (Cited on page 1.)

Pires, S., Vandenbussche, V., Kansal, V., Bender, R., Blot, L., Bonino, D., Boucaud, A., Brinchmann, J., Capobianco, V., Carretero, J. et al., 2020. Euclid: Reconstruction of weak-lensing mass maps for non-gaussianity studies. A&A **638**:A141. URL https://doi.org/10.1051/0004-6361/201936865. (Cited on page 95.)

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo et al., 2020. Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics* **641**:A6. (Cited on page 4.)

Porth, L. and R. E. Smith, 2021. Fast estimation of aperture mass statistics II: Detectability of higher order statistics in current and future surveys. arXiv e-prints arXiv:2106.04594. (Cited on page 26.)

Porth, L., R. E. Smith, P. Simon, L. Marian and S. Hilbert, 2020. Fast estimation of aperture mass statistics – I. Aperture mass variance and an application to the CFHTLenS data. Monthly Notices of the Royal Astronomical Society **499**(2):2474–2492. URL https://doi.org/10.1093/mnras/staa2900. (Cited on pages 26, 80, 85, 88, 91, 97, 102, and 104.)

Pratten, G. and A. Lewis, 2016. Impact of post-Born lensing on the CMB. *Journal of Cosmology and Astro-Particle Physics* **2016**(8):047. (Cited on page 79.)

Rimes, C. D. and A. J. S. Hamilton, 2005. Information content of the non-linear matter power spectrum. *Monthly Notices of the Royal Astronomical Society* **360**:L82–L86. (Cited on page 142.)

Robertson, H. P., 1935. Kinematics and World-Structure. *The Astrophysical Journal* **82**:284. (Cited on page 2.)

Rugh, S. E. and H. Zinkernagel, 2010. Weyl's principle, cosmic time and quantum fundamentalism. (Cited on page 3.)

Sadeh, I., F. B. Abdalla and O. Lahav, 2016. ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning. **128**(968):104502. (Cited on page 19.)

Sato, M., M. Takada, T. Hamana and T. Matsubara, 2011. Simulations of Wide-field Weak-lensing Surveys. II. Covariance Matrix of Real-space Correlation Functions. *The Astrophysical Journal* **734**(2):76. (Cited on pages 31 and 79.)

Scherrer, R. J. and E. Bertschinger, 1991. Statistics of primordial density perturbations from discrete seed masses. *The Astrophysical Journal* **381**:349–360. (Cited on page 13.)

Schneider, P., 1996. Detection of (dark) matter concentrations via weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society* **283**(3):837–853. (Cited on pages 36, 37, 80, and 82.)

Schneider, P., 1998. Cosmic Shear and Biasing. *The Astrophysical Journal* **498**(1):43–47. (Cited on pages 80 and 117.)

Schneider, P., 2006a. Part 1: Introduction to gravitational lensing and cosmology. In G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek & J. Wambsganss, editors, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, pages 1–89. (Cited on page 81.)

Schneider, P., 2006b. Part 3: Weak gravitational lensing. In G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek & J. Wambsganss, editors, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, pages 269–451. (Cited on pages 31, 79, and 81.)

Schneider, P., T. Eifler and E. Krause, 2010. Cosebis: Extracting the full e-/b-mode information from cosmic shear correlation functions. A&A **520**:A116. URL https://doi.org/10.1051/0004-6361/201014235. (Cited on pages 33 and 80.)

Schneider, P. and M. Kilbinger, 2007a. The ring statistics - how to separate E- and B-modes of cosmic shear correlation functions on a finite interval. *Astronomy & Astrophysics* **462**(3):841–849. (Cited on pages 32 and 33.)

Schneider, P. and M. Kilbinger, 2007b. The ring statistics - how to separate E- and B-modes of cosmic shear correlation functions on a finite interval. *Astronomy & Astrophysics* **462**(3):841–849. (Cited on page 80.)

Schneider, P., M. Kilbinger and M. Lombardi, 2005. The three-point correlation function of cosmic shear. II. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures. *Astronomy & Astrophysics* **431**:9–25. (Cited on pages 80, 81, 83, and 100.)

Schneider, P. and M. Lombardi, 2003. The three-point correlation function of cosmic shear. I. The natural components. *Astronomy & Astrophysics* **397**:809–818. (Cited on pages 32 and 79.)

Schneider, P., L. van Waerbeke, B. Jain and G. Kruse, 1998. A new measure for cosmic shear. *Monthly Notices of the Royal Astronomical Society* **296**(4):873–892. (Cited on pages 18, 32, 33, 34, 37, 38, 40, 41, 44, 65, 68, 69, 78, 79, 80, 82, 84, 85, 88, 104, and 148.)

Schneider, P., L. van Waerbeke, M. Kilbinger and Y. Mellier, 2002a. Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *Astronomy & Astrophysics* **396**:1–19. (Cited on pages 33 and 80.)

Schneider, P., L. van Waerbeke and Y. Mellier, 2002b. B-modes in cosmic shear from source redshift clustering. *Astronomy & Astrophysics* **389**:729–741. (Cited on pages 32, 37, 39, 40, 79, 80, 82, and 94.)

Scoccimarro, R., 1998. Transients from initial conditions: a perturbative analysis. *Monthly Notices of the Royal Astronomical Society* **299**:1097–1118. (Cited on page 49.)

Scoccimarro, R. and J. Frieman, 1996. Loop Corrections in Nonlinear Cosmological Perturbation Theory. *The Astrophysical Journal Suppl.* **105**:37–+. (Cited on page 82.)

Sefusatti, E., M. Crocce, S. Pueblas and R. Scoccimarro, 2006. Cosmology and the bispectrum. *Physical Review D* **74**(2):023522. (Cited on page 31.)

Seitz, S., P. Schneider and J. Ehlers, 1994. Light propagation in arbitrary spacetimes and the gravitational lens approximation. Classical and Quantum Gravity **11**(9):2345–2373. (Cited on pages 34 and 78.)

Seljak, U. and M. Zaldarriaga, 1996. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *The Astrophysical Journal* **469**:437–+. (Cited on pages 8 and 49.)

Sellentin, E. and A. F. Heavens, 2017. Quantifying lost information due to covariance matrix estimation in parameter inference. *Monthly Notices of the Royal Astronomical Society* **464**(4):4658–4665. (Cited on page 63.)

Semboloni, E., H. Hoekstra, J. Schaye, M. P. van Daalen and I. G. McCarthy, 2011. Quantifying the effect of baryon physics on weak lensing tomography. *Monthly Notices of the Royal Astronomical Society* **417**:2020–2035. (Cited on pages 32 and 79.)

Silk, J., 1968. Cosmic Black-Body Radiation and Galaxy Formation. *The Astrophysical Journal* **151**:459. (Cited on page 9.)

Skilling, J., 2004. Nested sampling. AIP Conference Proceedings **735**(1):395–405. URL https://aip.scitation.org/doi/abs/10.1063/1.1835238. (Cited on page 22.)

Smith, R. E., 2009. Covariance of cross-correlations: towards efficient measures for large-scale structure. *Monthly Notices of the Royal Astronomical Society* pages 1337–+. (Cited on page 49.)

Smith, R. E. and R. E. Angulo, 2019. Precision modelling of the matter power spectrum in a Planck-like Universe. *Monthly Notices of the Royal Astronomical Society* **486**(1):1448–1479. (Cited on page 13.)

Smith, R. E. and L. Marian, 2015. Towards optimal estimation of the galaxy power spectrum. *Monthly Notices of the Royal Astronomical Society* **454**(2):1266–1289. (Cited on page 62.)

Smith, R. E., J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou and H. M. P. Couchman, 2003. Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society* **341**:1311–1332. (Cited on pages 13 and 97.)

Smith, R. E., D. S. Reed, D. Potter, L. Marian, M. Crocce and B. Moore, 2014. Precision cosmology in muddy waters: cosmological constraints and N-body codes. *Monthly Notices of the Royal Astronomical Society* **440**:249–268. (Cited on page 62.)

Springel, V., 2005. The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society* **364**:1105–1134. (Cited on page 49.)

Swendsen, R. H. and J.-S. Wang, 1986. Replica monte carlo simulation of spin-glasses. Phys. Rev. Lett. **57**:2607–2609. URL https://link.aps.org/doi/10.1103/PhysRevLett.57.2607. (Cited on page 22.)

Szapudi, I. and A. S. Szalay, 1997. Cumulant correlators from the apm. *The Astrophysical Journal* **481**(1):L1–L4. (Cited on page 84.)

Takada, M. and B. Jain, 2003. The Three-Point Correlation Function for Spin-2 Fields. *The Astrophysical Journal Lett.* **583**(2):L49–L52. (Cited on page 32.)

Takahashi, R., M. Sato, T. Nishimichi, A. Taruya and M. Oguri, 2012. Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *The Astrophysical Journal* **761**:152. (Cited on pages 13 and 97.)

Takahashi, R., N. Yoshida, M. Takada, T. Matsubara, N. Sugiyama, I. Kayo, T. Nishimichi, S. Saito and A. Taruya, 2011. Non-Gaussian Error Contribution to Likelihood Analysis of the Matter Power Spectrum. *The Astrophysical Journal* **726**:7. (Cited on page 62.)

Taylor, A., B. Joachimi and T. Kitching, 2013. Putting the precision in precision cosmology: How accurate should your data covariance matrix be? *Monthly Notices of the Royal Astronomical Society* **432**(3):1928–1946. (Cited on page 63.)

Taylor, A. N. and P. I. R. Watts, 2001. Parameter information from non-linear cosmological fields. *Monthly Notices of the Royal Astronomical Society* **328**:1027–1038. (Cited on page 31.)

Tegmark, M., A. N. Taylor and A. F. Heavens, 1997. Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets? *The Astrophysical Journal* **480**:22–+. (Cited on page 62.)

Troxel, M. A. and M. Ishak, 2015. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. Phys. Rep. **558**:1–59. (Cited on pages 31 and 79.)

Troxel, M. A., N. MacCrann, J. Zuntz, T. F. Eifler, E. Krause, S. Dodelson, D. Gruen, J. Blazek, O. Friedrich, S. Samuroff et al., 2018. Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear. *Physical Review D* **98**(4):043528. (Cited on pages 31 and 79.)

Turchin, V. F., 1971. On the computation of multidimensional integrals by the monte-carlo method. Theory of Probability & Its Applications **16**(4):720–724. URL https://doi.org/10.1137/1116083. (Cited on page 22.)

Van Horn, K. S., 2003. Constructing a logic of plausible inference: a guide to cox's theorem. International Journal of Approximate Reasoning **34**(1):3–24. URL https://www.sciencedirect.com/science/article/pii/S0888613X03000513. (Cited on page 20.)

Van Waerbeke, L., Y. Mellier, T. Erben, J. C. Cuilland re, F. Bernardeau, R. Maoli, E. Bertin, H. J. McCracken, O. Le Fèvre, B. Fort et al., 2000. Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. *Astronomy & Astrophysics* **358**:30–44. (Cited on pages 31 and 78.)

Villaescusa-Navarro, F., C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys et al., 2020. The Quijote Simulations. *The Astrophysical Journal Suppl.* **250**(1):2. (Cited on page 133.)

Vogelsberger, M., 2020. Cosmological simulations of galaxy formation. Nature Reviews Physics **2**(1):42–66. URL https://doi.org/10.1038/s42254-019-0127-2. (Cited on page 11.)

Wadadekar, Y., 2005. Estimating Photometric Redshifts Using Support Vector Machines. **117**(827):79–85. (Cited on page 19.)

Wald, R. M., 1984. General relativity. Chicago Univ. Press, Chicago, IL. URL https://cds.cern.ch/record/106274. (Cited on page 1.)

Walker, A. G., 1937. On milne's theory of world-structure*. Proceedings of the London Mathematical Society **s2-42**(1):90–127. URL https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-42.1.90. (Cited on page 2.)

Wang, H., H. J. Mo, X. Yang, Y. P. Jing and W. P. Lin, 2014. ELUCID—Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics. *The Astrophysical Journal* **794**(1):94. (Cited on page 130.)

Weinberg, S., 1972. Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. Wiley, New York, NY. URL https://cds.cern.ch/record/100595. (Cited on page 3.)

Weinberg, S., 2008. Cosmology. Cosmology, by Steven Weinberg. ISBN 978-0-19-852682-7. Published by Oxford University Press, Oxford, UK, 2008. (Cited on page 11.)

Weyl, H., 1923. Zur allgemeinen Relativitätstheorie. Physikalische Zeitschrift **23**:230–232. (Cited on page 3.)

White, M. and W. Hu, 2000. A New Algorithm for Computing Statistics of Weak Lensing by Large-Scale Structure. *The Astrophysical Journal* **537**(1):1–11. (Cited on page 17.)

Wick, G. C., 1950. The evaluation of the collision matrix. Phys. Rev. **80**:268–272. URL https://link.aps.org/doi/10.1103/PhysRev.80.268. (Cited on page 13.)

Wittman, D. M., J. A. Tyson, D. Kirkman, I. Dell'Antonio and G. Bernstein, 2000. Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. Nature (London) **405**(6783):143–148. (Cited on pages 31 and 78.)

Yoshida, H., 1990. Construction of higher order symplectic integrators. Physics Letters A **150**(5):262–268. URL https://www.sciencedirect.com/science/article/pii/0375960190900923. (Cited on page 25.)

Zeldovich, Y. B., 1972. A hypothesis, unifying the structure and the entropy of the Universe. *Monthly Notices of the Royal Astronomical Society* **160**:1P. (Cited on page 13.)

Zhang, P., M. Liguori, R. Bean and S. Dodelson, 2007. Probing Gravity at Cosmological Scales by Measurements which Test the Relationship between Gravitational Lensing and Matter Overdensity. *Physical Review Letters* **99**(14):141302. (Cited on pages 30 and 78.)